

Search for single top–quark production with neural networks using 2.7 fb^{-1}

Dominic Hirschbühl, Jan Lück, Thomas Müller, Adonis Papaikonomou,
Manuel Renz, Wolfgang Wagner

KIT, Universität Karlsruhe

Abstract

We report on a search for electroweak single top–quark production with CDF II data corresponding to 2.7 fb^{-1} of integrated luminosity. We apply neural networks to construct discriminants that distinguish between single–top–quark and background events. Two analyses are performed, assuming a top–quark mass of $175 \text{ GeV}/c^2$. In the first one, we combine t – and s –channel events to one single–top signal under the assumption that the ratio of the two processes is given by the standard model. Using ensemble tests we determine that we expect with a probability of 50% to see a single-top signal that is larger than a 5.1σ fluctuation of the background (p –value of 0.00000018). A binned likelihood fit to the data yields a cross section of $2.1^{+0.7}_{-0.6} \text{ pb}$ for single top–quark production. The observed p –value is 0.00006790 which corresponds to a significance of 3.8σ .

In the second analysis, we separate the two single top–quark production modes. A binned likelihood fit done simultaneously to a two–dimensional and three one–dimensional distributions of neural network outputs yields most probable values for the cross sections of $0.7^{+0.6}_{-0.5} \text{ pb}$ for the t –channel and $2.1^{+0.7}_{-0.7} \text{ pb}$ for the s –channel.

Contents

0	Additions since Blessing of 2.7 fb^{-1} result	3
1	Introduction	5
2	Requirements for Candidate Events	6
3	Functionality of NeuroBayes[®]	9
3.1	The Training Process	10
3.2	Preprocessing of the Variables	10
3.3	Automatic Variable Selection	11
3.4	Training Result	11
4	Data Modeling and Event Yield	11
4.1	Signal MC	12
4.2	Modeling of backgrounds	14
4.3	Expected Event Yield	15
5	Training of the Neural-Networks	15
5.1	Training Samples	16
5.2	Input Variables	17
5.2.1	Top-Quark Reconstruction	17
5.2.2	Relevance of Input Variables	20
5.2.3	Comparison between Observation and Simulation	27
5.3	Output Distributions and Templates	33
5.3.1	Combined Search	33
5.3.2	Separate Search	36
6	Systematic Uncertainties	42
7	Likelihood Function	59
8	Ensemble Tests	61
8.1	Expected Sensitivity	61
8.2	Expected Significance	61
9	Application to Observed Events	62
9.1	Comparison to Expectation	62
9.2	Fit Results	66
9.2.1	Combined Search	66
9.2.2	Separate Search	67
A	Input Variables	68
A.1	Variables for the 2jet 2tag bin neural network	68
A.2	Variables for the 3jet 1tag bin neural network	77
A.3	Variables for the 3jet 2tag bin neural network	90

0 Additions since Blessing of 2.7 fb^{-1} result

Since the last blessing (CDF note 9107v1.6) we added the following points:

1. We add high- p_T muon data coming from MET+jets trigger as described in CDF note 9105. This new category of events will be called Extended Muon Coverage (EMC), opposite to the Trigger Lepton Coverage (TLC).
2. We add about 0.5 fb^{-1} of data (p14 – p17). Figure 1 and 2 show the neural network output distributions of the new observed events compared to the old observed events in all 8 channels.

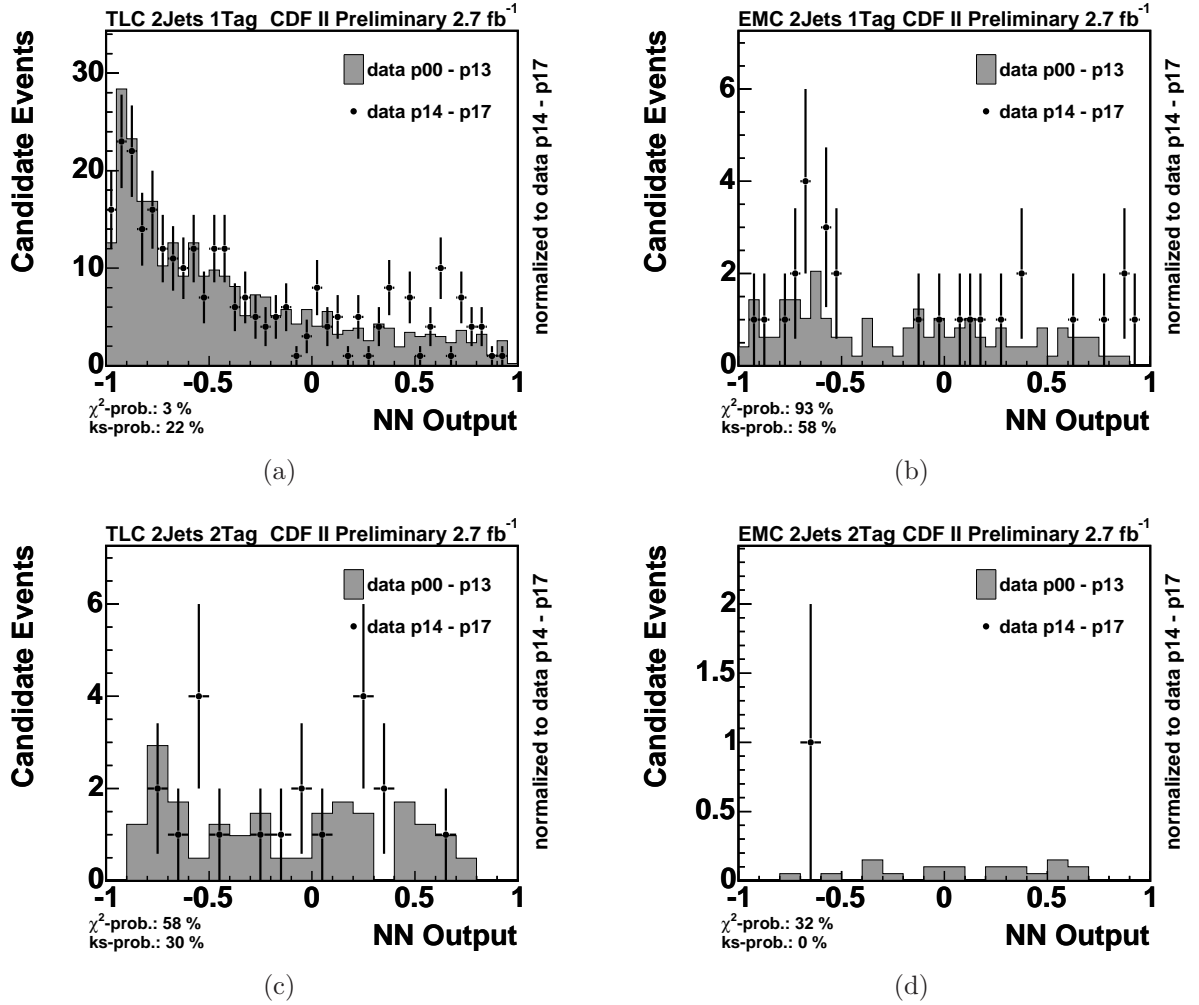
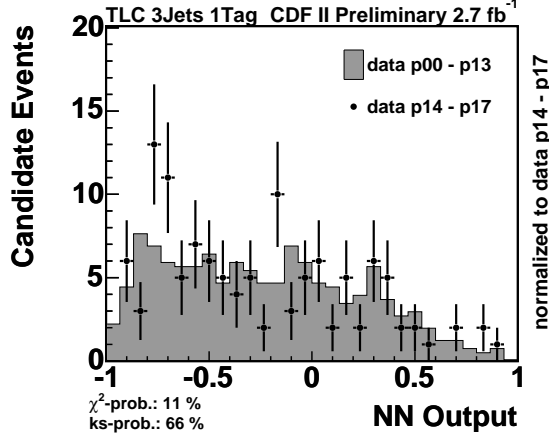
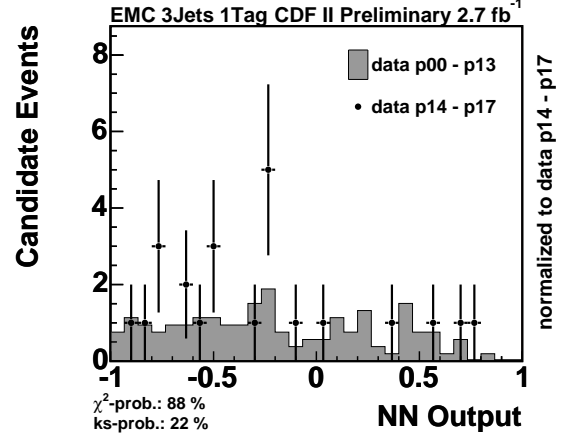


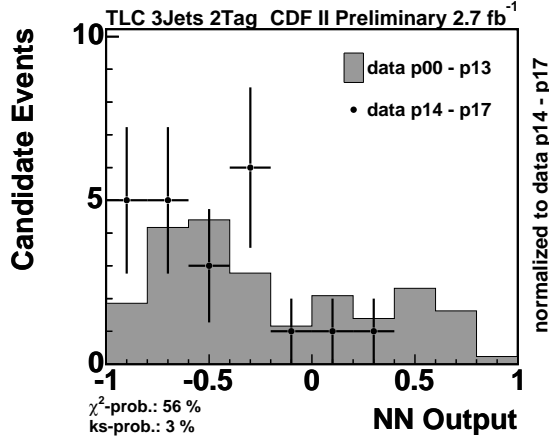
Figure 1: Comparison of NN Output distributions of the new observed events in the 2 jet 1 tag channel for TLC leptons (a) and EMC leptons (b); and in the 2 jet 2 tag channel for TLC leptons (c) and EMC leptons (d).



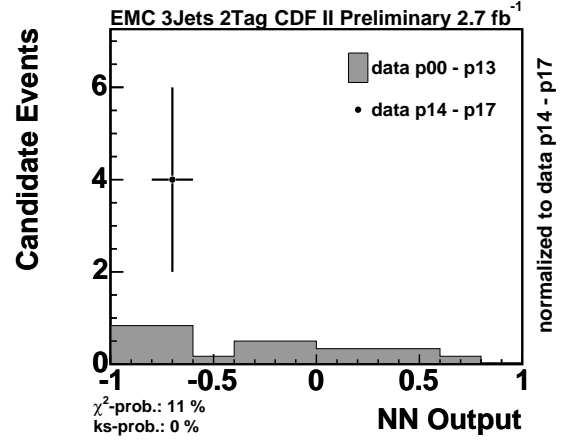
(a)



(b)



(c)



(d)

Figure 2: Comparison of NN Output distributions of the new observed events in the 3 jet 1 tag channel for TLC leptons (a) and EMC leptons (b); and in the 3 jet 2 tag channel for TLC leptons (c) and EMC leptons (d).

1 Introduction

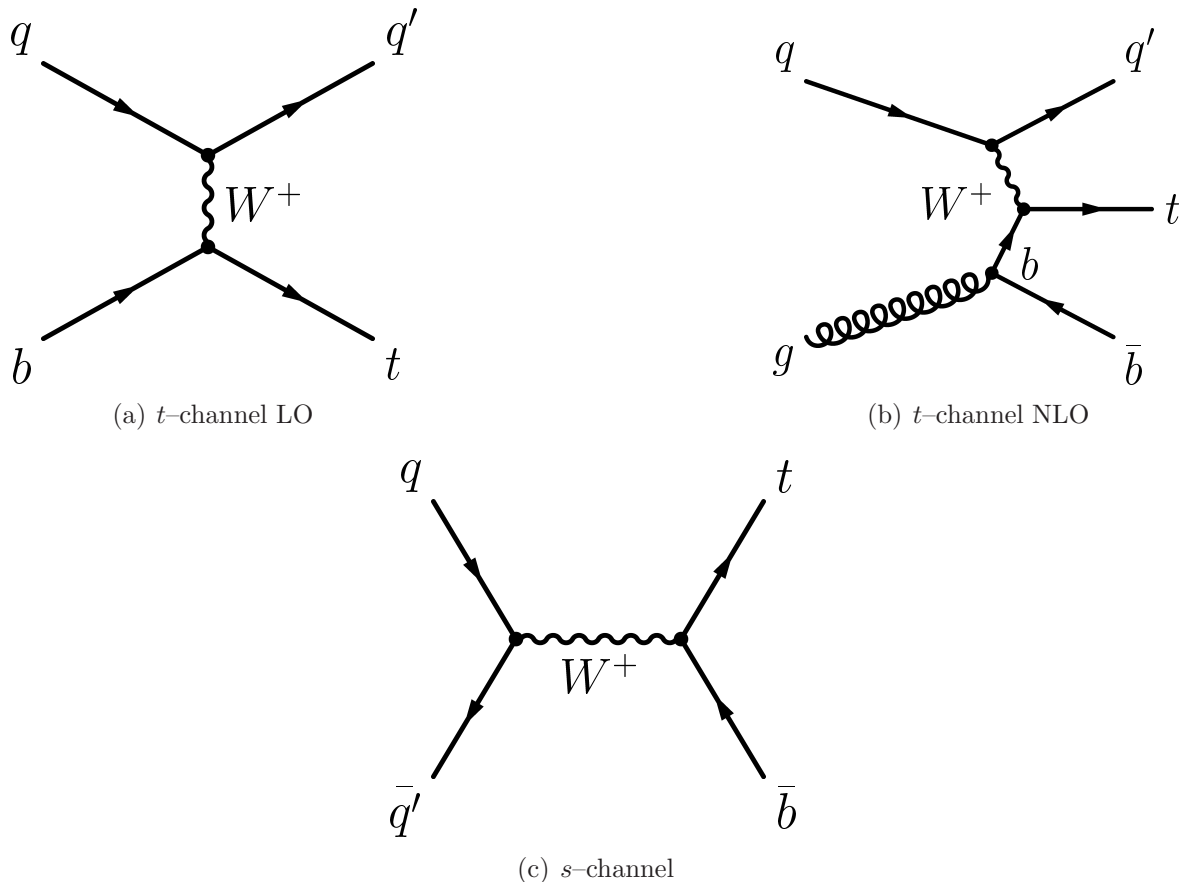


Figure 3: Single top-quark production channels: (a) t -channel at leading order, (b) t -channel NLO including initial state gluon splitting, and (c) s -channel at leading order. For antitop-quark production, the charge conjugate processes apply.

In this note, we document our analysis of CDF II data searching for electroweak production of single top-quarks. The two relevant processes of single top-quark production at the Tevatron are illustrated in figure 3. In this search, only leptonic W -boson decays in events with exactly two or three tight jets are targeted as described in section 2. Since a small signal must be extracted from a huge amount of background events, multivariate methods are an appropriate approach, exploiting a large number of characteristic quantities. In this analysis, neural networks are used to discriminate single-top-quark events by employing the NeuroBayes[®] package [?, ?] introduced in section 3. The modeling of the signal and background processes is described in section 4. In this section, a table with the number of expected events is given as well. The training of the neural networks is detailed in section 5. To measure the cross section of single top-quark production, signal and background templates of the neural network output are created from simulated events, which are fitted to the output distribution of observed events.

For the search of single-top-quark events, two different approaches are taken. One is

to assume the ratio of the cross sections of both production channels to be as predicted by the standard model. In this approach, the output distributions of t - and s -channel events are combined into one single template. The other is to determine the cross sections of both channels independently and simultaneously, named separate search.

We use the high- p_T lepton trigger data up to p13. The largest data sample (CEM and CMUP) corresponds to an integrated luminosity of 2.7 fb^{-1} . To exclude runs or run sections where the CDF II detector was not fully operational, we apply the single-top group version 19 of the good run list (bits 1141).

2 Requirements for Candidate Events

Lepton Identification After offline reconstruction, the lepton candidates have to pass further cuts in order to improve the purity. For central electrons, a reconstructed track with $p_T > 10 \text{ GeV}/c$ has to match a cluster in the CEM with $E_T > 20 \text{ GeV}$. Furthermore, $E_{\text{HAD}}/E_{\text{EM}} < 0.055 + 0.00045 \cdot E$ is required; the ratio of cluster energy to track momentum E/p has to be smaller than 2.0 for track momenta $\leq 50 \text{ GeV}/c$. Electron candidates in forward direction are defined by a cluster in the PEM with $E_T > 20 \text{ GeV}$ and $E_{\text{HAD}}/E_{\text{EM}} < 0.05$. The cluster position and the primary vertex are combined to form a search trajectory in the silicon tracker and seed the pattern recognition of the tracking algorithm. For the electron candidate to pass the selection, the found track has to fulfill certain quality criteria. Electron events are rejected, if an additional high- p_T track is found that forms a common vertex with the track of the electron candidate and has a curvature of opposite sign. These events are likely to stem from the conversion of a photon.

Muons are identified by requiring a COT track with $p_T > 20 \text{ GeV}/c$ that extrapolates to a track segment in a muon chamber. Signal muons have to be detected in the CMU and CMP simultaneously or in the CMX. In order to minimize background contaminations further requirements are imposed. The energy depositions in the electromagnetic and hadronic calorimeters have to correspond to the expectation regarding minimum ionizing particles. To reject cosmic muons or muons from in-flight decays of long-lived particles (such as K_S, K_L , or Λ), the impact parameter d_0 of the track must be small. Cosmic muons are further rejected through their characteristic track timing and topology.

Furthermore, exactly one isolated lepton candidate is required, whereby a candidate is considered isolated if the E_T not assigned to the lepton in a cone of $R = 0.4$ centered around the lepton is less than 10% of the lepton E_T or p_T , respectively. This lepton is called tight lepton.

A detailed description of all lepton requirements can be found in references [?, ?, ?].

Dilepton Veto To ensure that there is exactly one tight lepton, events are rejected which have either an additional tight lepton or a loose lepton. Loose leptons are leptons which pass all cuts except the isolation cut, or are identified in the CMP, CMU or BMU solely.

Jet Reconstruction and Selection In this analysis, jets are reconstructed with a cone of $R = 0.4$ without taking into account calorimeter towers which are associated to any tight isolated electron. The jet energy is corrected up to level 5, i.e. for the η -dependence of the calorimeter response, for multiple $p\bar{p}$ interactions, and absolute energy scale (i.e. up to the underlying particle jet). Candidate jets, required to have detector $|\eta| < 2.8$, must have corrected $E_T > 20$ GeV to be called tight jets, whereas loose jets must have corrected E_T between 12 GeV and 20 GeV. Detector η is defined as the pseudorapidity of the jet calculated with respect to the origin of the coordinate system, which is located in the center of the detector. Only events with exactly two or three tight jets are accepted, whereby at exactly one or two of the jets must be tagged as a b -quark jet by requiring a displaced secondary vertex within the jet.

Missing Transverse Energy The missing E_T ($\vec{\cancel{E}}_T$) is defined by

$$\vec{\cancel{E}}_T = - \sum_i E_T^i \hat{n}_i, \quad (1)$$

where i denotes the calorimeter tower number with $|\eta| < 3.6$, \hat{n}_i is a unit vector perpendicular to the beam axis which points at the i^{th} calorimeter tower. Additionally, $\cancel{E}_T = |\vec{\cancel{E}}_T|$ is defined. Because this calculation is based on calorimeter towers, $\vec{\cancel{E}}_T$ has to be adjusted for the effect of the jet corrections for all tight and loose jets.

Since muons pass the calorimeters without showering, i.e. as minimum ionizing particle, a correction is applied by adding all transverse momenta of the traversing muons to the sum and by removing the average ionization energy. The corrected \cancel{E}_T is required to be greater than 25 GeV.

Rejection of QCD multi-jet background To further suppress events in which no real W boson is produced, additional cuts are applied. The cuts are based on the assumption that these events do not produce \cancel{E}_T by nature but due to lost or mismeasured jets. Therefore, one would expect small \cancel{E}_T , small \cancel{E}_T significance $\cancel{E}_{T \text{ sig}}$, a small transverse W -boson mass $M_{T,W}$, and small values of the angle $\Delta\phi_{\vec{\cancel{E}}_T, \text{jet}}$ between $\vec{\cancel{E}}_T$ and a jet. The \cancel{E}_T significance is defined as

$$\cancel{E}_{T \text{ sig}} = \frac{\cancel{E}_T}{\sqrt{\sum_{\text{jets}} C_{\text{JES}}^2 \cos^2(\Delta\phi_{\vec{\cancel{E}}_T, \text{jet}}) + \cos^2(\Delta\phi_{\vec{\cancel{E}}_T^{\text{uncorr}}, \vec{\cancel{E}}_T^{\text{corr}})}}} \quad (2)$$

with the level 5 jet-correction factor C_{JES} and the azimuthal angle $\Delta\phi_{\vec{\cancel{E}}_T^{\text{uncorr}}, \vec{\cancel{E}}_T^{\text{corr}}}$ between uncorrected and corrected $\vec{\cancel{E}}_T$. The transverse W -boson mass is given by

$$M_{T,W} = \sqrt{2p_T^\ell \cancel{E}_T - \vec{p}_T^\ell \cdot \vec{\cancel{E}}_T} \quad (3)$$

with p_T^ℓ being the transverse momentum of the charged lepton.

Events passing the electron trigger must have $M_{T,W} > 20$ GeV. For central electrons, it is additionally required that $\cancel{E}_{T \text{ sig}} > -0.05M_{T,W} + 3.5$. In the 2- and 3-jet bin,

$\cancel{E}_{T \text{ sig}} > 2.5 - 2.5|\Delta\phi_{\vec{\cancel{E}}_T, \text{jet2}}|/0.8$ must be fulfilled, with the azimuthal angle $\Delta\phi_{\vec{\cancel{E}}_T, \text{jet2}}$ between $\vec{\cancel{E}}_T$ and the second leading jet. In the 1-jet bin, $\cancel{E}_{T \text{ sig}} > -7.6 + 3.2|\Delta\phi_{\ell, \text{jet}}|/0.8$ must be satisfied, with $\Delta\phi_{\ell, \text{jet}}$ being the azimuthal angle between the charged lepton and the jet. The CEM QCD veto in the 2-jet bin is exemplarily illustrated in figure 4. Forward electrons must have $\cancel{E}_{T \text{ sig}} > 2.0$. Additionally, $\cancel{E}_T > 45 - 30|\Delta\phi_{\vec{\cancel{E}}_T, \text{jet}}|$ must be

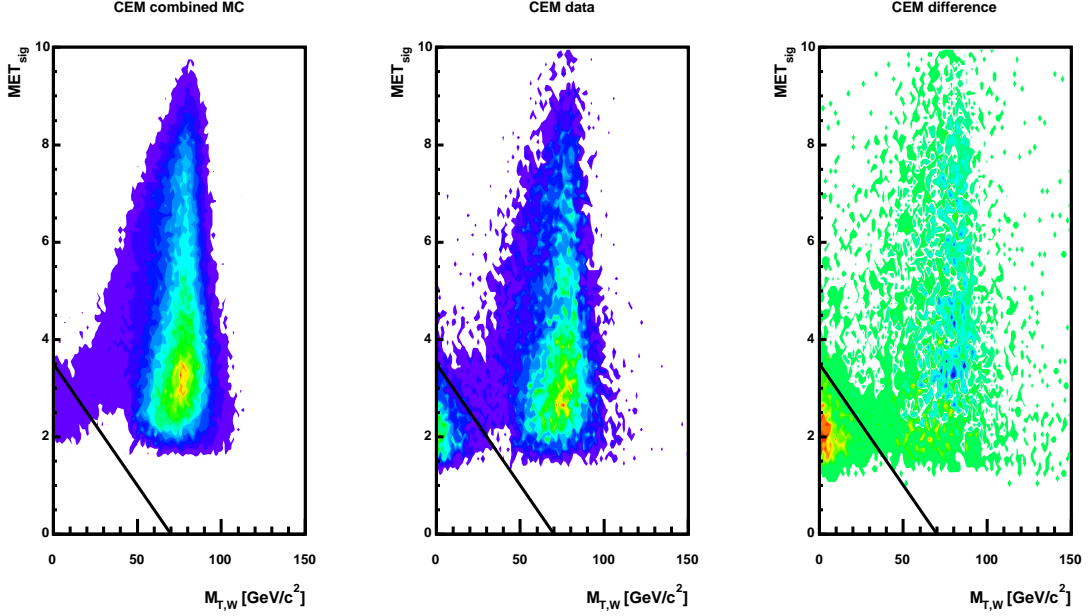


Figure 4: Illustration of the CEM QCD veto in the 2-jet bin. The distribution of $\cancel{E}_{T \text{ sig}}$ versus $M_{T,W}$ is shown in the pretag sample. On the left hand side, the distribution of the W +jets sample is shown. The plot in the middle shows the distribution of pretag data. On the right hand side, the difference between the other two distributions is shown. The line represents the cut $\cancel{E}_{T \text{ sig}} > -0.05M_{T,W} + 3.5$. This cut rejects events which are not modeled by the W +jets sample.

fulfilled for the two leading jets in each event.

In muon events, the transverse W -boson mass must fulfill $M_{T,W} > 10$ GeV. In the 1-jet bin, events containing CMUP muons are additionally required to have $\cancel{E}_T > 60|\Delta\phi_{\ell, \text{jet}}| - 145$.

Z -Boson Veto To remove Z -boson events, events are rejected in which the charged lepton can be paired with any more loosely defined jet or lepton to form an invariant mass consistent with the Z peak, defined as the range from $76 \text{ GeV}/c^2$ to $106 \text{ GeV}/c^2$.

3 Functionality of NeuroBayes[®]

NeuroBayes[®] combines a three-layer feed-forward neural network with a complex robust preprocessing. This preprocessing is performed before the input variables are fed to the neural network. The neural network uses Bayesian regularization techniques for the training process. The network infrastructure consists of one input node for each input variable plus one bias node, an arbitrary number of hidden nodes, and one output node which gives a continuous output in the interval $[-1, 1]$.

The nodes of two consecutive layers are catenated with variable connections. For each node j , a biased weighted sum of the values of the previous layer x_i is calculated

$$a_j(\mathbf{x}) = \sum_i \omega_{ij} x_i + \mu_{0,j} \quad (4)$$

and passed to the transfer function which gives the output of the node. The bias $\mu_{0,j}$ implements the threshold of node j . The output of each node is determined by a transformed sigmoid function

$$S(\mathbf{x}) = \frac{2}{1 + e^{-a(\mathbf{x})}} - 1 \quad (5)$$

which gives an output of -1 for background and $+1$ for signal. As can be seen in figure 5, the sigmoid function is only sensitive to a relatively small range around zero. By this transformation, the interval $[-\infty, +\infty]$ is mapped to the interval $[-1, +1]$. For very large ($x \rightarrow \infty$) or very small ($x \rightarrow -\infty$) values, a saturation effect is reached. The bias mentioned above shifts the mean of the sum of the weighted input data distribution $\sum_i \omega_{ij} x_i$ to the linear part of the sigmoid function.

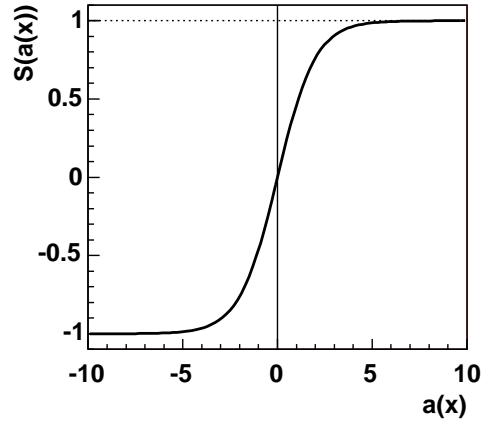


Figure 5: The transformed sigmoid activation function $S(a(x))$ as given by equation 5.

The output of the neural network for the output node is calculated by

$$o = S\left(\sum_{j=0}^M \omega_j \cdot S\left(\sum_{i=0}^d \omega_{ij} x_i + \mu_{0,j}\right)\right) \quad (6)$$

where d is the number of input nodes and M the number of hidden nodes. ω_{ij} denotes the weights from the input to the hidden layer, ω_j the weights from the hidden layer to the output node. $\mu_{0,j}$ is the weight that connects the bias node with the hidden nodes.

3.1 The Training Process

The training of a neural network is done by minimizing the deviation between the true output and the one calculated by using the actual weights. The error function minimized in this neural network is the entropy error function

$$E_D = \sum_i \log\left(\frac{1}{2} \cdot (1 + T_i) \cdot o_i + \epsilon\right), \quad (7)$$

where the target value T_i is a binary number to classify event i as signal or background, o_i represents the output as given by equation (6). ϵ is a small regularization constant which is introduced in order to avoid numerical problems at the beginning of the training. This constant is reduced in each training iteration and is zero after just a few iterations.

The aim of the training of a neural network is to find the minimum in the multidimensional structure of the error function which may exhibit many peaks and valleys. As this task can be difficult to solve, the training process is done by the combined method of backpropagation and gradient descent, i.e. the change of each weight $\Delta\omega_{ij}$ is adjusted proportional to the current gradient of the error function $\Delta\omega_{ij} = -\eta \frac{\partial E_D}{\partial \omega_{ij}}$. The step width η is adapted individually for each weight during the training. Since the target value is not known for hidden nodes, the error induced by the current weights has to be propagated backwards from the output node by applying the chain rule for partial derivatives.

The neural network is trained with regularization techniques to improve generalization performance and to avoid overtraining. During the training process, the weights are systematically reduced in addition to the variation calculated by the gradient descent procedure. Thus, only recurring structures are intensified while the influence of statistical fluctuations is reduced by so-called weight decay. Connections (and even nodes) that have become completely insignificant are pruned away. This reduces the number of free parameters and hence improves the signal-to-noise ratio by removing the cause of the noise, leading to an improved generalization ability. For details of the above mentioned features see references [?, ?].

3.2 Preprocessing of the Variables

To find the optimal starting point for minimizing the error function, the input variables are preprocessed. This preprocessing is done in a completely automatic way. Equalizing the input variables and scaling them to be distributed between -1 and 1 before passing the variables to the neural network reduces the influence of extreme outliers. Those flattened distributions are then converted into Gaussian distributions, centered at zero with standard deviation one. At the beginning of the training, this avoids saturation of the nodes due to the above mentioned shape of the activation function (see figure 5) and assures that also the inputs to the next layers are distributed with mean zero and width one. To decorrelate the preprocessed input variables, at first, their covariance matrix is

calculated. Diagonalizing the covariance matrix using Jacobi rotations [?] and dividing the rotated input vectors by the square root of the corresponding eigenvalue transforms the covariance matrix into a unit matrix.

The above mentioned transformation to a Gaussian distribution may be altered by individual variable preprocessing like fitting a spline curve to the flattened distribution. In addition, discrete variables can be treated as members of classes. The preprocessing of those kinds of variables can also deal with a certain order of values, e.g. the number of tracks in a jet. The preprocessing is also able to deal with variables that are only given for a subset of events by assigning the missing values to a δ function.

3.3 Automatic Variable Selection

The significances of the training variables are determined automatically during the preprocessing in NeuroBayes[®].

The correlation matrix of all preprocessed input variables is calculated including the correlation of all variables to the target. One variable after the other is omitted to determine the loss of total correlation to the target caused by its removal. The variable with the smallest loss of correlation is discarded leading to an $(n - 1)$ -dimensional correlation matrix. The same procedure is repeated with the reduced correlation matrix to find the least important of the $(n - 1)$ remaining variables.

The significance of each variable is calculated by dividing the loss of correlation induced by its removal at the relevant point of the successive procedure by the square root of the sample size, i.e. those significances are relative numbers in terms of the reduced correlation matrices.

After the preprocessing process, it is possible to cut on the significance of the variables to incorporate only those that include relevant information that is not already incorporated by other variables. The number of discarded variables is determined by scanning the sorted list, starting with the least relevant one, until the first quantity has a significance larger than the required minimum value.

3.4 Training Result

As already mentioned above, the network output of signal events piles up at $+1$, while background events accumulate at outputs around -1 . This is illustrated in figure 6(a). After minimizing the entropy error function (7), the output, rescaled to the interval $[0, 1]$, can be interpreted as Bayesian *a posteriori* probability, if the *a priori* probability is correct, i.e. if a realistic mixture of signal and background has been chosen. Hence, the quality of the training can be checked by plotting the signal purity for each output bin, as illustrated in figure 6(b). If the network is well trained, all the points should lie on the diagonal.

For a detailed discussion of the output interpretation, see reference [?].

4 Data Modeling and Event Yield

Using multivariate methods, it is crucial to model the observed data correctly. Therefore, the complete spectrum of expected processes contributing to the considered W +jets bins

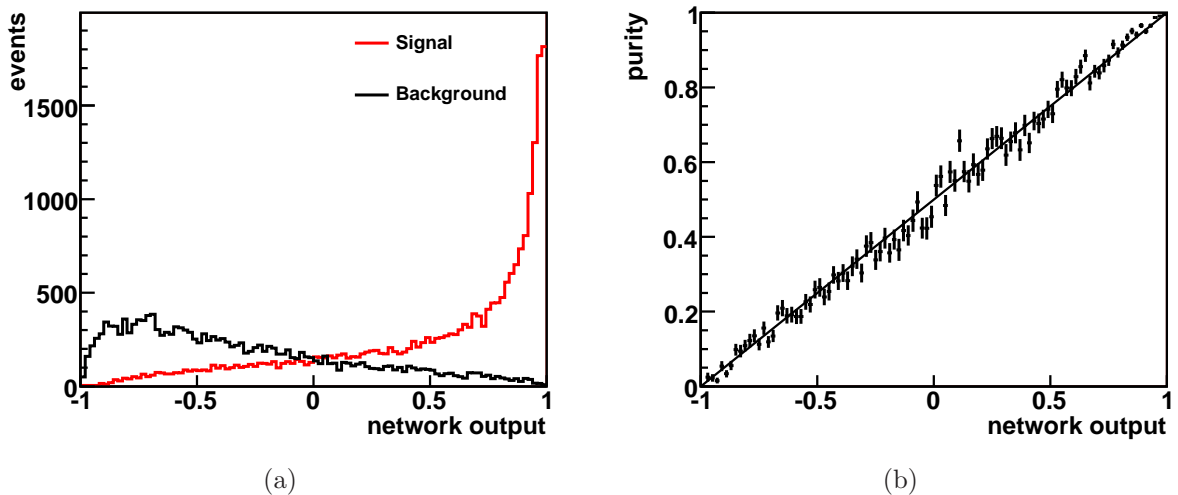


Figure 6: Illustration of the training result. Figure (a) shows the output distributions for signal and background events. Figure (b) shows the signal purity of each output bin and illustrates the expected linear dependence.

has to be modeled. Most of the processes are described using Monte Carlo simulation, while some background processes are derived from data.

4.1 Signal MC

Several authors pointed out [?, ?] that the leading order contribution to single top-quark production via t -channel, as modeled in leading order parton shower Monte Carlo programs, does not adequately represent the measured final states.

The leading order process is a $2 \rightarrow 2$ process with a b quark in the initial state as given by figure 3(a): $b + u \rightarrow d + t$ or $b + \bar{d} \rightarrow \bar{u} + t$. For antitop-quark production, the charge conjugate processes are implied. The initial-state b -quark being part of the quark sea is described by a b -quark PDF for the calculation. Since the b quark originally stems from a gluon splitting into a $b\bar{b}$ pair, a \bar{b} quark has to be present in the event. Leading order parton shower programs create this \bar{b} quark through backward evolution following the DGLAP scheme [?, ?, ?]. Thereby, only the soft regime of the transverse momentum distribution of the \bar{b} quark is modeled well, while the high- p_T tail is not estimated adequately. Also, the pseudorapidity spectrum expands too far into the forward region.

One can improve the modeling of single top-quark production via t -channel by producing two samples of simulated events with matrix element generators and applying a parton shower Monte Carlo program to the final-state partons. In this analysis, the matrix element generator MADEVENT, interfaced to the CTEQ5L [?] parameterization of the parton distribution functions (PDF), was used to produce simulated events. Parton showering and hadronization were performed using PYTHIA.

The first sample is the $2 \rightarrow 2$ process $b + q \rightarrow q' + t$ given by the Feynman graph in figure 3(a); the second process is a $2 \rightarrow 3$ process with a gluon in the initial state,

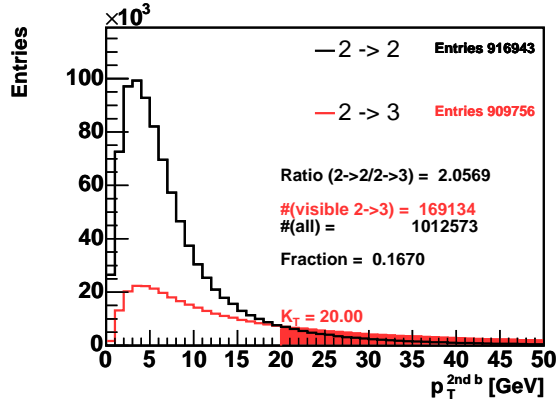


Figure 7: Matching of single-top-quark events produced by the $2 \rightarrow 2$ and the $2 \rightarrow 3$ t -channel processes. The p_T distributions of the 2nd b quark in the event are shown. The ratio of $2 \rightarrow 2$ to $2 \rightarrow 3$ events is adjusted such that the rate of 2nd b quarks with $p_T > 20 \text{ GeV}/c$ and $|\eta| < 2.8$ matches the NLO prediction. The fraction of these events is illustrated by the shaded area.

$g + q \rightarrow q' + t + \bar{b}$, which is shown in figure 3(b). In the second process, the \bar{b} quark, called 2nd b quark in the following, is produced directly in the hard scattering described by the matrix element. This sample describes the most important next-to-leading order (NLO) contribution to t -channel production and is therefore suitable to describe the high- p_T tail of the p_T distribution of the 2nd b quark.

However, the two samples, the $2 \rightarrow 2$ process and the $2 \rightarrow 3$ process, have to be matched to form one unified sample of simulated events. This has been realized by a procedure of adjusting the ratio between the $2 \rightarrow 2$ and $2 \rightarrow 3$ processes in such a way that the rate of events with a detectable 2nd b -quark jet, that is $p_T > 20 \text{ GeV}/c$ and $|\eta| < 2.8$, matches the prediction made by ZTOP [?], a program which operates at NLO in perturbation theory. The matched t -channel sample is created in such a manner that the p_T distribution of 2nd b -quarks in the matched t -channel sample consists of $2 \rightarrow 2$ events for transverse momenta below a certain cutoff and of $2 \rightarrow 3$ events for transverse momenta above the cutoff. This cutoff and hence the ratio between the two processes are varied until the rate of detectable 2nd b -quark jets matches the prediction [?]. The final ratio is found to be $R = 2.1$, the cutoff is derived to be $K_T = 20 \text{ GeV}/c$, as displayed in figure 7. As a result, all detectable 2nd b quarks ($p_T > 20 \text{ GeV}/c$) in the matched t -channel sample are provided by the $2 \rightarrow 3$ process. The 2nd- b -quark p_T spectrum in the matched t -channel sample is much harder than the one provided by the $2 \rightarrow 2$ process alone.

More details on the t -channel matching procedure and the comparison to ZTOP can be found in reference [?].

4.2 Modeling of backgrounds

For the modeling of $t\bar{t}$, diboson, simulated events generated with PYTHIA are used, while the W +heavy flavor and Z +heavy flavor backgrounds were simulated using a combination of ALPGEN and PYTHIA. For the samples produced with ALPGEN, it is important that the heavy flavor overlap is removed. In this analysis, this is done by the jet-based procedure.

The QCD background is modeled using two different approaches [?]. The central electron and muon models are obtained from central electron trigger data. The events are required to pass all kinematic electron cuts but to fail two of the five non-kinematic cuts¹. Even though their kinematic properties resemble those of W -like events, those events are QCD-enriched, since the non-kinematic criteria serve primarily to filter out QCD-induced multijet events. For the forward electron sample, such a model is not yet available. For this reason, an additional QCD model is introduced, based on the idea that for a QCD event to pass the selection criteria, a jet has to resemble an electron. Hence, events from jet trigger data are required to have a jet with $E_T > 20$ GeV, $0.05 < E_{\text{HAD}}/E_{\text{EM}} < 0.2$, and at least four reconstructed tracks. The latter makes it unlikely that the event contains a real electron. Once a jet is identified as a fake electron, its charge is assigned randomly, and it is further considered as a tight forward electron.

The event candidates of both approaches have to pass all but the lepton selection criteria and the b -tag requirement to contribute to the corresponding QCD model. Since demanding a tagged jet would cause too low statistics, the b tag in the event has to be faked using taggable² jets. If, in a given event, only one jet is taggable, this one is considered to be the tagged jet. If there are more taggable jets in the event, one of those jets is randomly assigned to be tagged. In doing so, each taggable jet has the same probability to be selected.

Due to the usage of the neural-network b -tagger described in reference [?], it is additionally necessary to assign a hypothesis of what kind of quark flavor the jet is, b , c , or light-quark flavor. The probability that a specific quark flavor is assigned to a certain jet is given by the expected flavor composition of the QCD background. This composition is estimated by applying the neural-network b -tagger to the $\cancel{E}_T < 15$ GeV sideband of the observed data. In this sideband sample, a flavor composition of 45% b -quark jets, 40% c -quark jets, and 15% light-quark jets is found [?].

In order to describe events with mistagged light-quark jets, W +light flavor and Z +light flavor events simulated with ALPGEN and showered with PYTHIA are used. Due to the very small fraction of tagged events in this sample, a large amount of this kind of events would be needed in simulation. Thus, the pretag sample is utilized where taggable jets are assigned to be tagged. If an event has more than one taggable jet, the probability that a specific jet is selected as the tagged one is given by the ratios of the mistag probabilities. This probability is defined by the negative tag rate and the correction factor for the mistag asymmetry. Additionally, each event is weighted by the mistag probability of the jet considered as tagged.

For both QCD and mistag model, the output of the neural network b tagger is randomly assigned to the jet attributed as tagged. For this purpose template output distributions

¹ $Q \cdot \Delta x, |\Delta z|, E_{\text{HAD}}/E_{\text{EM}}, L_{\text{shr}}, \chi_{\text{strip}}^2$

² $E_T > 10$ GeV, $|\eta| < 2.4$, and $N_{\text{trk}} \geq 2$.

obtained from jets of simulated events corresponding to the respective quark flavor are utilized.

For the W +jet and Z +jet events produced by ALPGEN, it is important to take into account that the same n -jet configuration can be generated starting from different $(n-m)$ -parton configurations, where the additional m partons are provided by the shower. Hence, to avoid double-counting of certain parts of the phase space, this necessitates a matching of the diverse parton configurations generated by the matrix element generator.

4.3 Expected Event Yield

The number of expected events, displayed in table 1, has been determined by using “Method2 for You”.

Process	Number of Events			
	2-jet-bin		3-jet-bin	
	1-tag-bin	2-tag-bin	1-tag-bin	2-tag-bin
$t\bar{t}$ dilepton	30.3 ± 4.3	8.1 ± 1.3	23.8 ± 3.3	7.8 ± 1.3
$t\bar{t}$ non-dilepton	62.8 ± 8.9	12.9 ± 2.0	177.3 ± 24.8	47.8 ± 7.8
$Wb\bar{b}$	366.5 ± 110.4	47.3 ± 14.7	104.1 ± 31.4	16.4 ± 5.1
$Wc\bar{c}/Wc$	340.3 ± 104.9	4.3 ± 1.5	87.0 ± 26.8	2.2 ± 0.8
$Wq\bar{q}$ Mistags	303.7 ± 38.0	1.3 ± 0.3	87.5 ± 11.1	0.9 ± 0.2
QCD multijets	55.9 ± 22.4	1.5 ± 0.6	21.5 ± 8.6	0.2 ± 0.1
Diboson	50.8 ± 5.0	3.1 ± 0.3	16.0 ± 1.6	1.0 ± 0.1
Z +jets	18.4 ± 2.7	0.8 ± 0.1	6.9 ± 1.0	0.5 ± 0.1
t -channel	49.8 ± 7.3	1.4 ± 0.2	12.9 ± 1.9	2.0 ± 0.3
s -channel	26.1 ± 3.7	7.3 ± 1.2	8.2 ± 1.1	2.5 ± 0.4
total prediction	1304.6 ± 220.2	86.8 ± 16.6	545.2 ± 65.7	81.3 ± 10.7
observation	1312	82	491	95

Table 1: Summary of predicted numbers of signal and background events in the selected data sample. All systematic uncertainties are included (see section 6).

5 Training of the Neural-Networks

In the search for single top-quark production, six different neural networks are trained in the different jet and tag bins:

- 2 jets, 1 tag: t -channel for combined search, t -channel for separate search, and s -channel for separate search
- 2 jets, 2 tags: s -channel for combined and separate search
- 3 jets, 1 tag: t -channel for combined and separate search

Process	t -channel	s -channel	t -channel	t -channel
	2 jets; 1 b -tag	2 jets; 2 b -tags	3 jets; 1 b -tag	3 jets; 2 b -tags
t -channel	50.0%	0.0%	50%	50%
s -channel	0.0%	50.0%	0.0%	0.0%
$t\bar{t}$	5.1%	15.9%	22.0%	37.0%
$Wb\bar{b}$	13.7%	27.2%	7.9%	11.2%
$Wc\bar{c}/Wc$	14.0%	3.2%	7.8%	1.8%
mistags	14.0%	0.0%	10.0%	—
Diboson	2.4%	2.2%	1.6%	—
Z +jets	0.8%	0.6%	0.7%	—

Table 2: Composition of the training samples used to train the neural networks to discriminate single-top-quark events for the combined search. The contributions of the respective single-top-quark events is 50% of the complete training sample. The relative fractions to the background category are given by table 1. The absolute fractions are determined by the requirement that the contributions of all background processes sum up to 50%.

- 3 jets, 2 tags: t -channel for combined and separate search

As introduced in section 3, the number of input nodes is given by the number of input variables which will be discussed in section 5.2.2. The number of hidden nodes was chosen to be 15.

5.1 Training Samples

For the training of the networks, it is necessary to arrange training samples consisting of the relevant modeled physics processes, each with reasonable statistics. A natural approach would be to select a mixture corresponding to the estimated composition of the observed events. Since this is not practicable in case of the search for single top-quark production, a different composition of the training samples has to be chosen.

The samples used to train the neural networks are composed in such a way that the respective signal single-top-quark production-process contributes 50% to the total number of events. The relative fractions of all considered background processes are given by the respective number of expected events quoted in table 1. For the spararate search one further needs rather good separation between s - and t -channel single top-quark events. Therefore, the training of those networks are special is the way, that the s -channel single top-quark sample is used as background sample for the training of the t -channel network and vice versa. Since the expected events for single-top-quark production are very low, we use five times the expected events for the mixture of the training composition.

A list of all used processes, modeled as described in section 4, as well as their contributions to the training samples is given in table 2 for the combined search and in table 3 for the separate search.

Process	t -channel	s -channel	s -channel	t -channel	t -channel
	2 jets 1 tag	2 jets 1 tag	2 jets 2 tags	3 jets 1 tag	3 jets 2 tags
t -channel	50.0%	9.6%	0.0%	50.0%	0.0%
s -channel	5.1%	50.0%	50.0%	0.0%	50.0%
$t\bar{t}$	4.6%	4.1%	15.9%	22.0%	37.0%
$Wb\bar{b}$	12.3%	11.1%	27.2%	7.9%	11.2%
$Wc\bar{c}/Wc$	12.6%	12.3%	3.2%	7.8%	1.8%
mistags	12.5%	11.3%	0.0%	10.0%	—
Diboson	2.2%	1.9%	2.2%	1.6%	—
Z +jets	0.7%	0.7%	0.6%	0.7%	—

Table 3: Composition of the training samples used to train the neural networks to discriminate single-top-quark events for the separate search. The contributions of the respective single-top-quark events is 50% of the complete training sample. The relative fractions to the background category are given by table 1. The absolute fractions are determined by the requirement that the contributions of all background processes sum up to 50%.

5.2 Input Variables

Before the training, the input variables are preprocessed as discussed in section 3.2 and the respective correlations to the target are determined, providing an ordered list of input variables being presented in section 5.2.2. In this connection, only variables with a significance larger than a certain threshold are utilized for the training.

For all neural-network trainings mentioned above, three categories of input variables are used: some are directly measured in the detector, others are reconstructed out of measured properties, and a few are calculated by advanced algorithms like the neural-network b -tagger described in reference [?]. Furthermore, some quantities need the knowledge about the four-vector of the top quark whose reconstruction is described in section 5.2.1.

5.2.1 Top-Quark Reconstruction

For some of the variables, the reconstruction of the top quark is necessary. The top-quark four-momentum is built out of the reconstructed W boson and a defined b -quark jet.

The first step in top-quark reconstruction is the selection of the b -quark jet. For the t -channel networks, the tagged jet is taken as the b -quark jet from the top-quark decay. For the s -channel neural networks, the jet with the largest product of the charge of the tight lepton (Q_ℓ) and the jet pseudorapidity, $Q_\ell \cdot \eta$, is assigned to belong to the top-quark decay.

The second step is the reconstruction of the four-momentum of the W boson, being built from the four-momenta of the measured tight lepton and the reconstructed neutrino. For the reconstruction of the neutrino, the smallest $|p_z^\nu|$ solution is chosen.

Rank	Variable	Relative Significance (in σ)
1	$M_{\ell\nu b}$	114
2	NN b -tag output (b from top)	86
3	M_{j1j2}	74
4	$Q_\ell \cdot \eta_j$	50
5	$M_{T,\ell\nu b}$	29
6	$\cos \Theta(\ell, lj)$	27
7	$E_{T,lj}$	19
8	$\cos \Theta(\ell, W)$	17
9	η_W	16
10	$M_{T,W}$	10
11	$\sum_{j1j2}(\eta_{jets})$	7
12	$p_{T,\ell}$	5
13	H_T	5
14	$\cos \Theta(\ell, W)_{\text{lab}}$	5

Table 4: Set of discriminating variables used for the training of the t -channel neural network in the 2-jet bin with 1 b -tag. The quoted relative significances are determined as described in section 3.3, i.e. calculated in terms of reduced matrices.

Rank	Variable	Relative Significance (in σ)
1	NN b -tag output (b from top)	100
2	$M_{\ell\nu b}$	76
3	H_T	46
4	$\cos \Theta(j, j)$	36
5	$\cos \Theta(\ell_{Wr.f.}, W)$	27
6	E_T^{lj}	23
7	$M_{T,W}$	19
8	$M_{T\ell\nu b}$	15
9	η_W	13
10	M_{j1j2}	9
11	$p_{T\ell}$	8
12	$\cos \Theta(W, \ell\nu b)$	6
13	$\cos \Theta(W, \ell\nu b)_{r.f.(top)}$	7
14	$Q_\ell \cdot \eta_j$	7
15	$\cos \Theta^*(\ell, W)$	7
16	$\cos \Theta(lep, beam)$	7
17	η_ℓ	6

Table 5: Set of discriminating variables used for the training of the s -channel neural network in the 2-jet bin with 1 b -tag. The quoted relative significances are determined as described in section 3.3, i.e. calculated in terms of reduced matrices.

Rank	Variable	Relative Significance (in σ)
1	$M_{\ell\nu j1j2}$	34
2	$M_{T,W}$	24
3	$M_{T,\ell\nu b}$	18
4	$\cos \Theta(j, j)$	18
5	$M_{\ell\nu b}$	11
6	$\sum_{\text{jets}}(\text{NN } b\text{-tag output})$	8
7	η_W	7
8	\cancel{E}_T (level 5)	6
9	M_{j1j2}	6
10	$E_{T,b}$	5
11	η_ℓ	3

Table 6: Set of discriminating variables used for the training of the s -channel neural network in the 2-jet bin with 2 b -tags. The quoted relative significances are determined as described in section 3.3, i.e. calculated in terms of reduced matrices.

Rank	Variable	Relative Significance (in σ)
1	$Q_\ell \cdot \eta_j$	38
2	NN b -tag output (b from top)	35
3	H_T	25
4	M_{j1j3}	23
5	$M_{\ell\nu b}$	19
6	$p_{T,\ell\nu b,jj}$	17
7	M_{j2j3}	13
8	$\cos \Theta(\ell, lj)$	12
9	$\Delta\eta(j1, j2)$	11
10	$\sum_{\text{jet}}(\eta_{jet})$	7
11	$\cancel{E}_T^{\text{sig}}$	6
12	M_{j1j2j3}	7
13	$E_{T,j2} + E_{T,j3}$	7
14	$E_{T,b}$	6
15	$M_{T,\ell\nu b}$	6
16	$\Delta\eta(t, lj)$	5
17	$E_{T,j1} + E_{T,j3}$	4
18	$E_{T,j1} + E_{T,j2}$	5

Table 7: Set of discriminating variables used for the training of the t -channel neural network in the 3-jet bin with 1 b -tag. The quoted relative significances are determined as described in section 3.3, i.e. calculated in terms of reduced matrices.

Rank	Variable	Relative Significance (in σ)
1	$Q_\ell \cdot \eta_{lj}$	25
2	$M_{\ell\nu bb}$	11
3	$p_{T,\ell\nu bjj}$	9
4	M_{j1j2}	6
5	$\cos \Theta(\ell, lj)$	5
6	$\sum_{j1j2}(E_{T,jets})$	5
7	M_{j1j3}	4
8	$\Delta\eta(j2, j3)$	4
9	$E_{T,2nd-b}$	4
10	$M_{T,\ell\nu b}$	3
11	$\Delta\eta(j1, j2)$	3
12	$E_{T,j3}$	2
13	centrality $\sum_{j1j2}(E_{T,jets})/\hat{s}$	2
14	\hat{s}	3
15	$\cos \Theta(j, j)$	3

Table 8: Set of discriminating variables used for the training of the t -channel neural network in the 3-jet bin with 2 b -tags. The quoted relative significances are determined as described in section 3.3, i.e. calculated in terms of reduced matrices.

5.2.2 Relevance of Input Variables

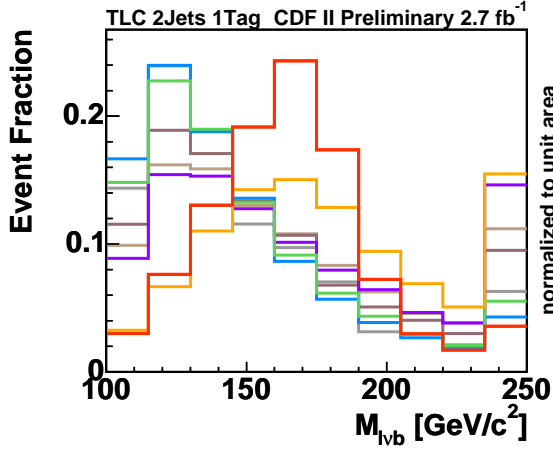
Many variables have initially been investigated for the training of the neural networks. Since the relevance of the variables varies for the five trainings, each neural network utilizes a different set of input variables passing the cut on the significance, quoted in tables 4, 5, 6, and 7.

Some of the utilized variables are:

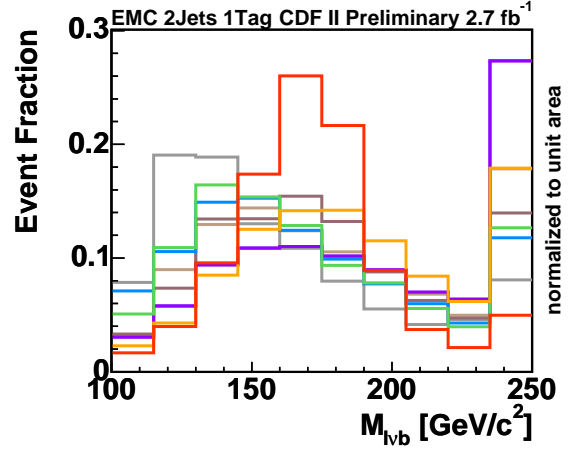
- the reconstructed top-quark mass $M_{\ell\nu b}$
- the output of the neural-network b -tagger of the selected b -quark jet from the top-quark decay
- the product of the charge of the lepton and the pseudorapidity of the light-quark jet $Q_\ell \cdot \eta_{lj}$
- the invariant mass of the two most energetic jets M_{j1j2}
- the invariant mass of the three most energetic jets M_{j1j2j3}
- the transverse mass of the reconstructed top quark $M_{T\ell\nu b}$
- the pseudorapidity of the reconstructed W boson η_W
- the transverse energy of the light-quark jet E_T^{lj}

- the cosine of the polar angle between the charged lepton in the W -boson rest-frame and the direction of the W boson
- the transverse W -boson mass $M_{T,W}$ as given by equation (3)
- the scalar sum of transverse energies $H_T = \sum_{\text{jets}} E_T + p_T^\ell + \cancel{E}_T$ where $\sum_{\text{jets}} E_T$ is the sum of the transverse energies of the tight jets
- the transverse energy of the selected b -quark jet from the top-quark decay E_T^b
- the transverse momentum of the combination of the reconstructed top quark and all additional jets $p_{T\ell\nu bj}$ and accordingly $p_{T\ell\nu bjj}$
- the transverse energy of the jet with the smallest transverse energy in the 3-jet bin E_T^{j3}
- the cosine of the polar angle between the reconstructed W boson in the top-quark rest-frame and the direction of the top quark $\cos\Theta(W, t)$
- the cosine of the polar angle between the reconstructed W boson and the direction of the top quark $\cos\Theta(W, t)_{\text{lab}}$
- the transverse momentum of the reconstructed top quark $p_{T\ell\nu b}$
- the transverse momentum of the charged lepton p_T^ℓ
- the pseudorapidity of the charged lepton η_ℓ
- the cosine of the angle in the top-quark rest-frame between the tight lepton and the beam axis $\cos\Theta(\ell, \text{beam})$
- the cosine of the angle in the top-quark rest-frame between the tight lepton and the light-quark jet $\cos\Theta(\ell, lj)$
- the sum of the outputs of the neural-network b -tagger of all tagged jets $\sum_{\text{jets}} (\text{NN } b\text{-tag output})$

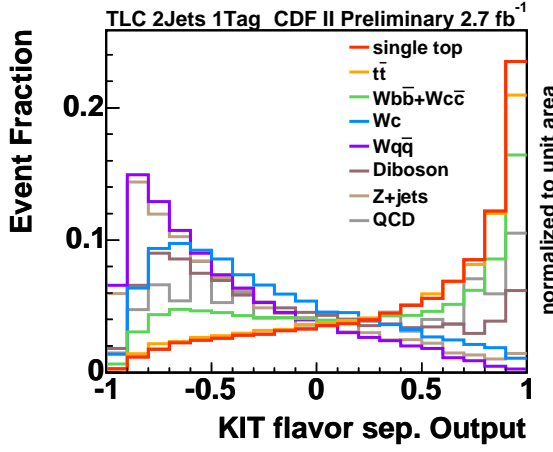
The distributions of the input variables of the t -channel network in the 2-jet bin with 1 b -tag for TLC and EMC leptons can be found in figures 8 to 12. The distributions of the s -channel network in the 2-jet bin with 1 and 2 b -tag, the t -channel network in the 3-jet bin with 1 and 2 b -tag can be found in the appendix.



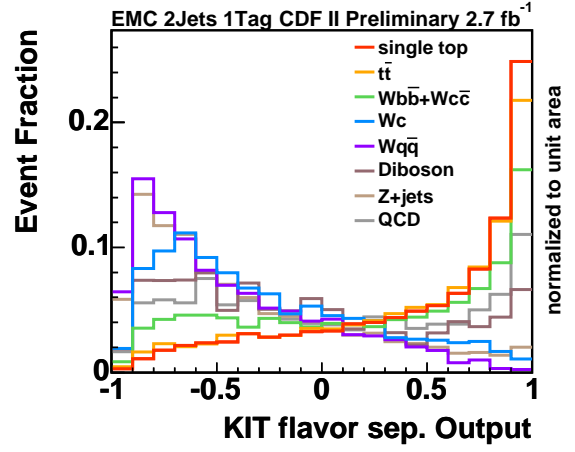
(a) $M_{\ell\nu b}$



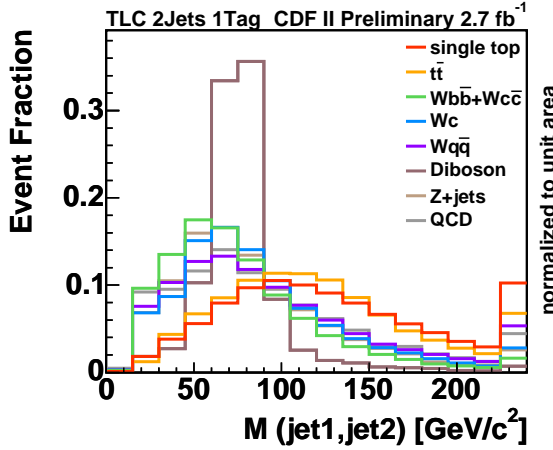
(b) $M_{\ell\nu b}$



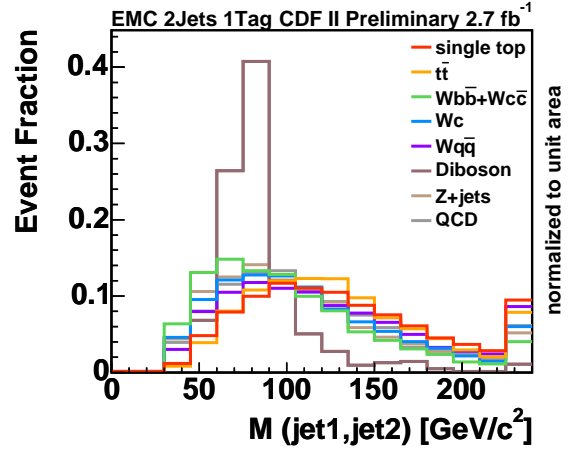
(c) NN b -tag output



(d) NN b -tag output



(e) M_{j1j2}



(f) M_{j1j2}

Figure 8: The three most important input variables of the t -channel network in the 2-jet bin with 1 b -tag for TLC (left) and EMC (right) leptons.

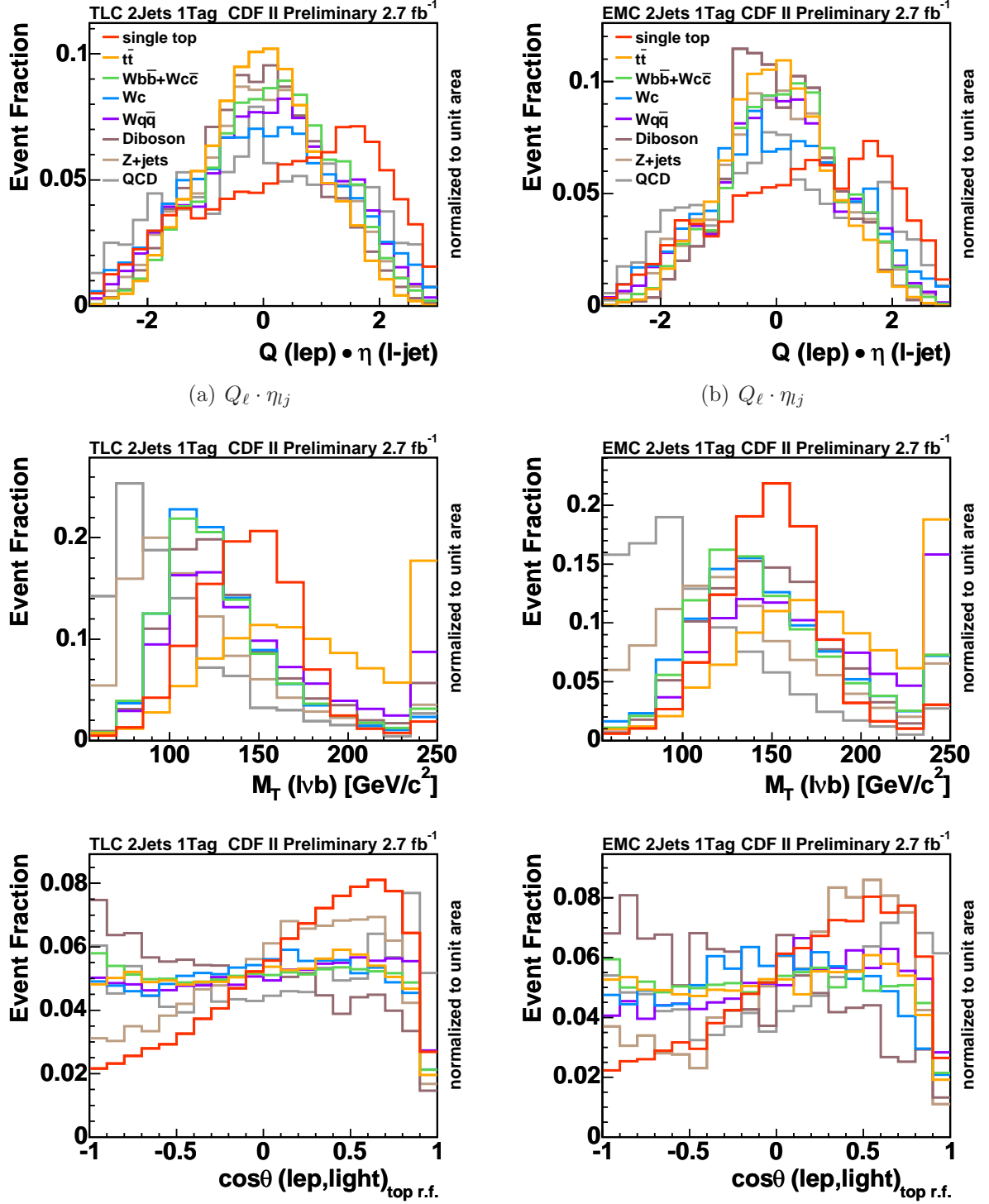


Figure 9: The input variables four to six of the t -channel network in the 2-jet bin with 1 b -tag for TLC (left) and EMC (right) leptons.

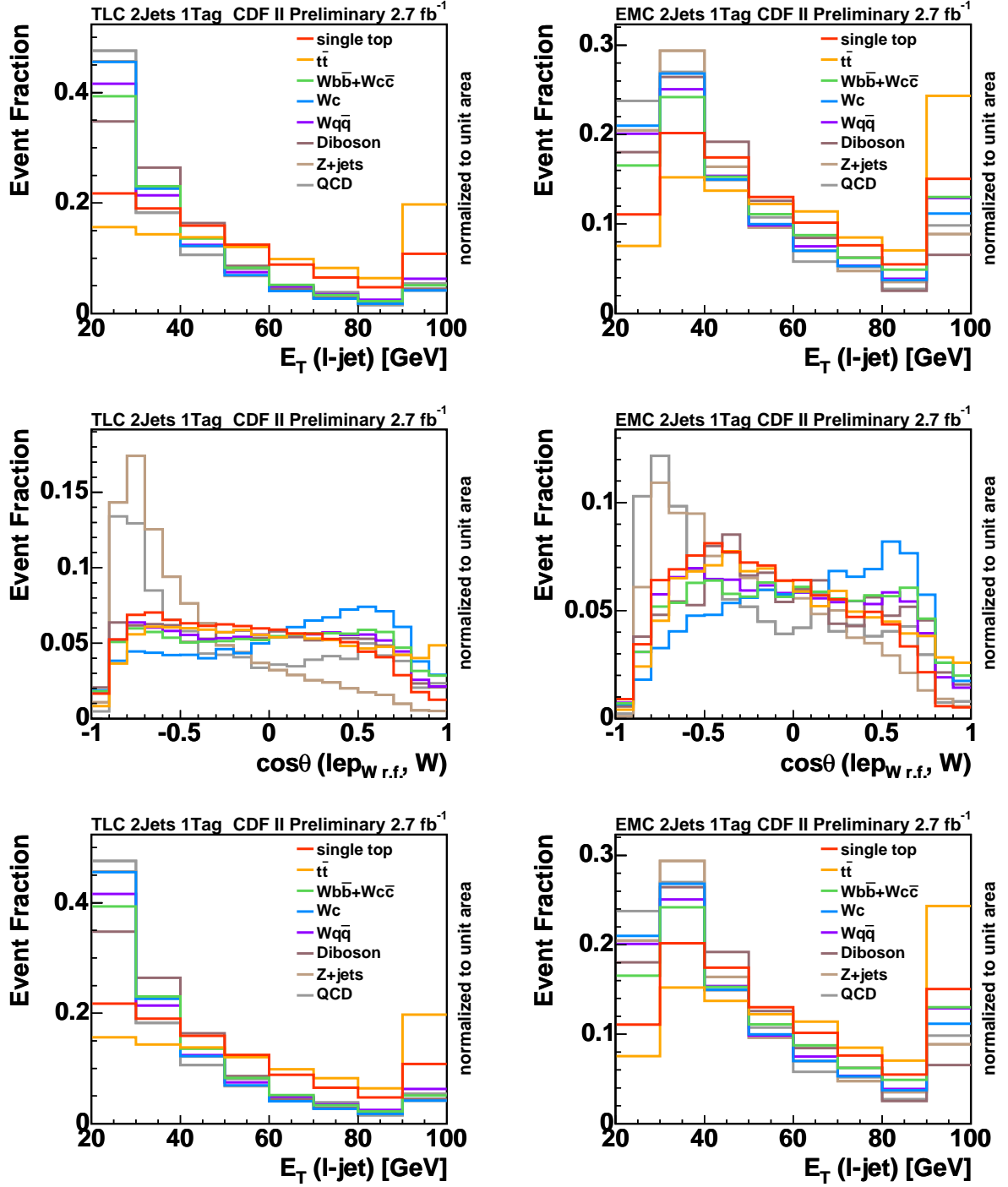


Figure 10: The input variables seven to nine of the t -channel network in the 2-jet bin with 1 b -tag for TLC (left) and EMC (right) leptons.

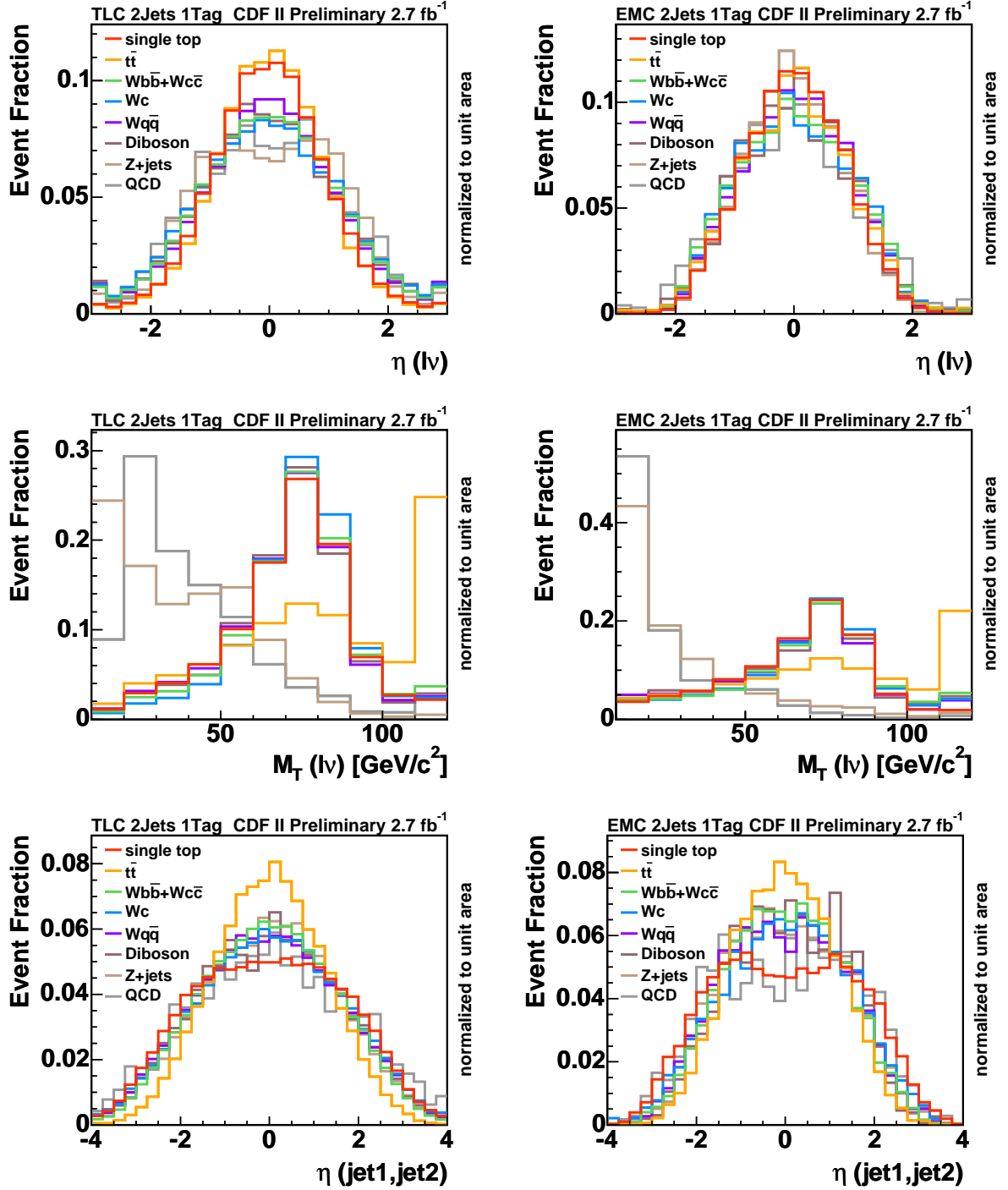


Figure 11: The input variables ten to twelve of the t -channel network in the 2-jet bin with 1 b -tag for TLC (left) and EMC (right) leptons.

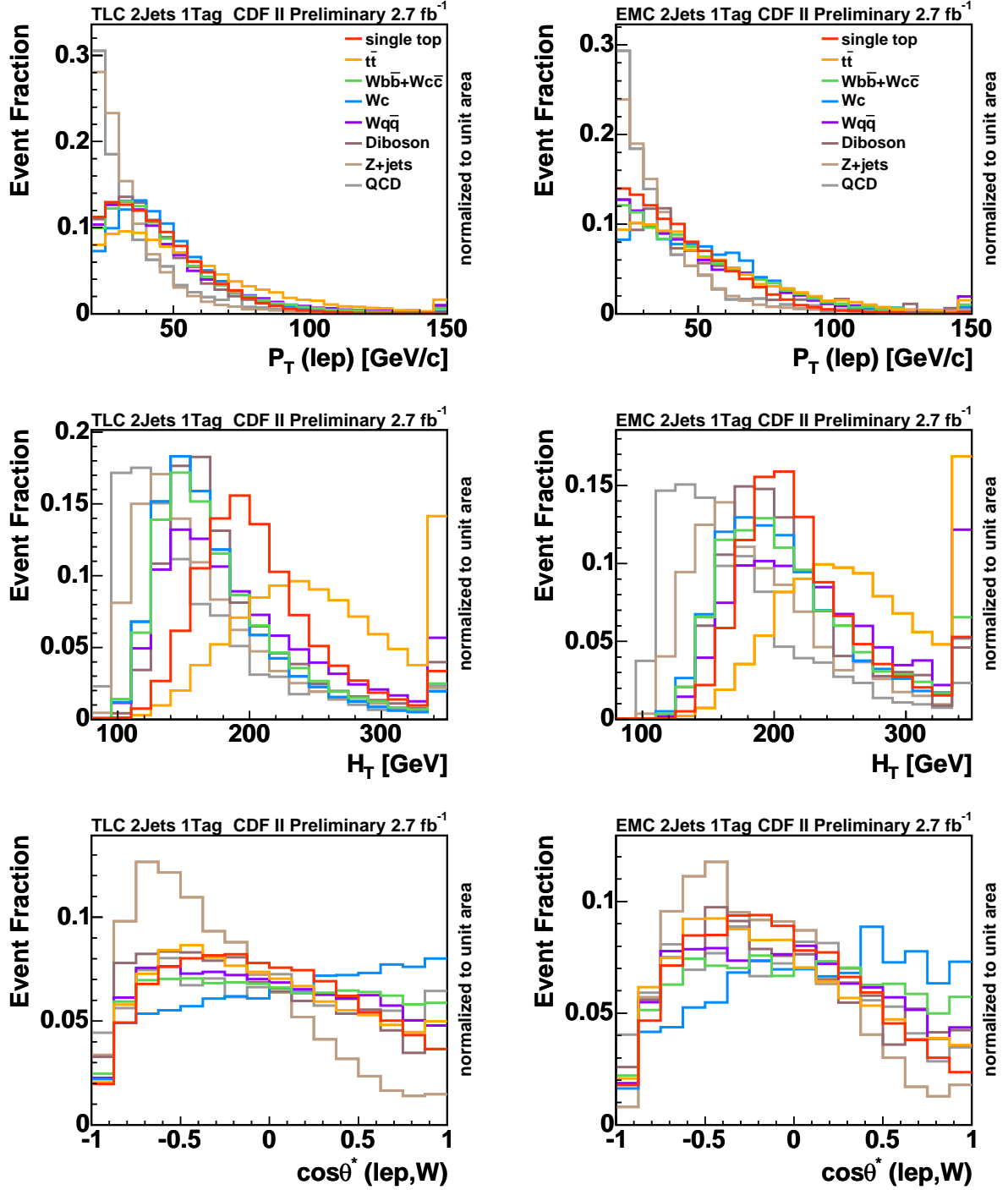


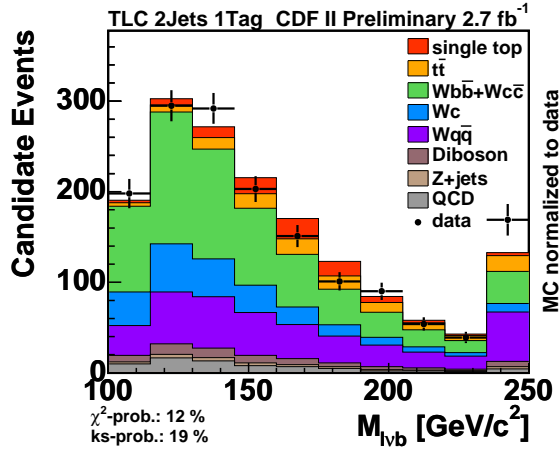
Figure 12: The input variables thirteen to fifteen of the t -channel network in the 2-jet bin with 1 b -tag for TLC (left) and EMC (right) leptons.

5.2.3 Comparison between Observation and Simulation

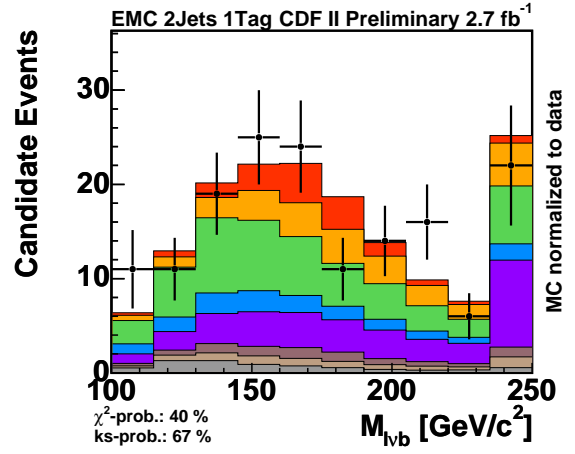
Since the neural network is trained with simulated events, it is crucial to check if the input variables are modeled correctly. Hence it is necessary to compare the shape of each input variable in observed events with the shape obtained by modeled data embracing the signal and background models described in section 4. For this comparison, each modeled process is scaled in such a way that it contributes as many events to the compound model as predicted by the estimated event yield quoted in table 1.

The comparisons of the distributions of the input variables of the t -channel network in the 2-jet bin with 1 b -tag can be found in figures 13 to 17. The distributions of the s -channel network in the 2-jet bin with 1 and 2 b -tag, the t -channel network in the 3-jet bin with 1 and 2 b -tag can be found in the appendix.

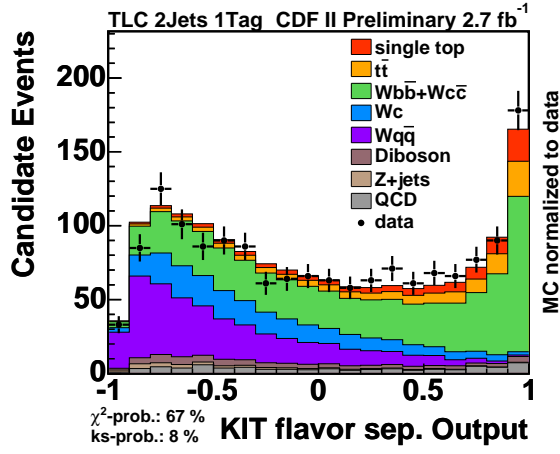
The distributions of the s -channel network in the 2-jet bin with 1 and 2 b -tag, the t -channel network in the 3-jet bin with 1 and 2 b -tag can be found in the appendix.



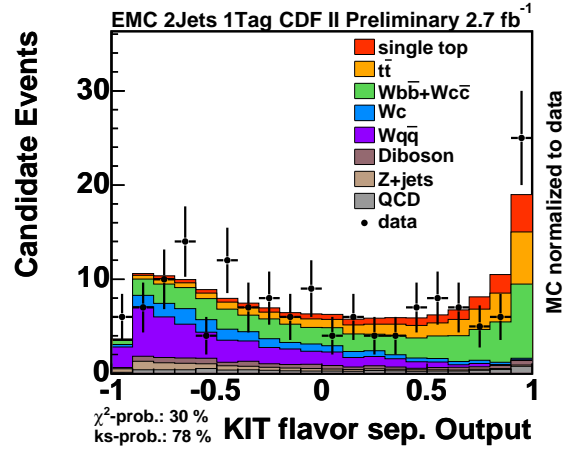
(a) $M_{\ell\nu b}$



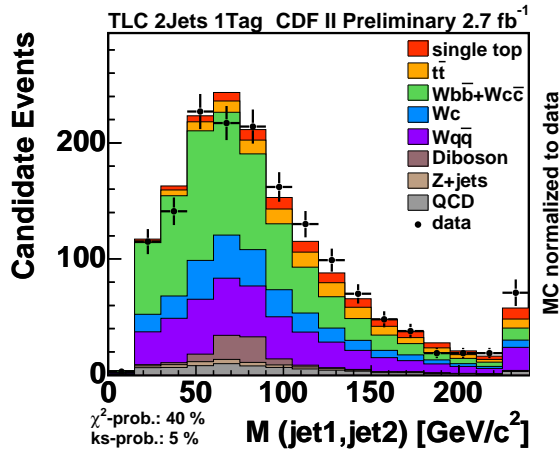
(b) $M_{\ell\nu b}$



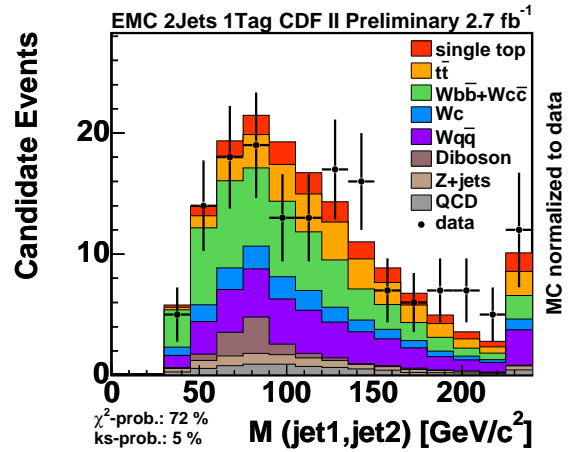
(c) NN b -tag output



(d) NN b -tag output

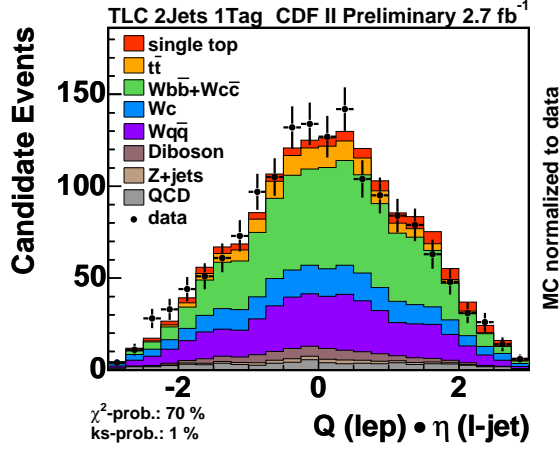


(e) M_{j1j2}

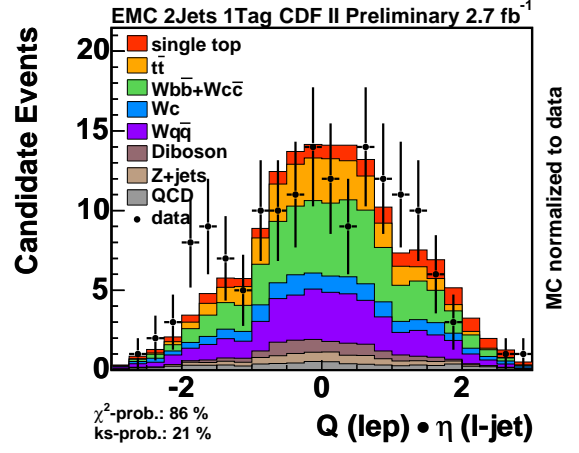


(f) M_{j1j2}

Figure 13: Data–MC comparison of the three most important input variables of the t -channel network in the 2-jet bin with 1 b -tag with 1 b -tag for TLC (left) and EMC (right) leptons.



(a) $Q_\ell \cdot \eta_j$



(b) $Q_\ell \cdot \eta_j$

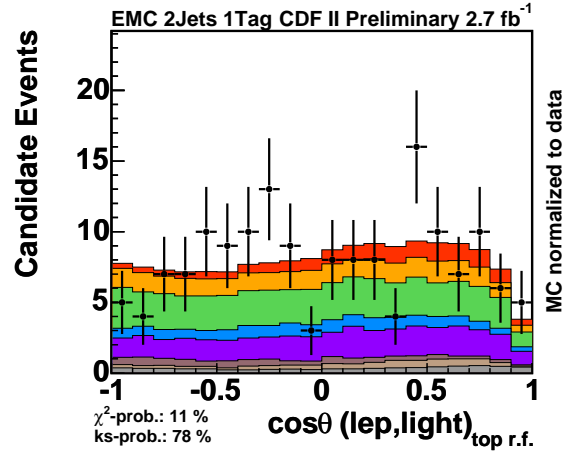
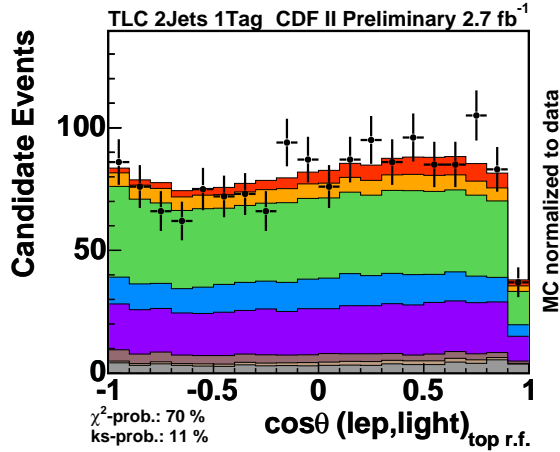
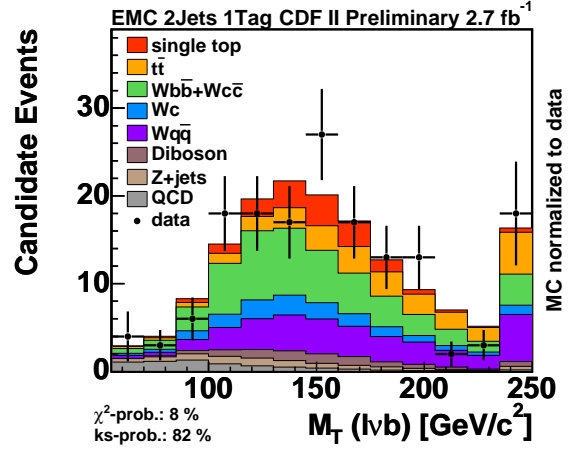
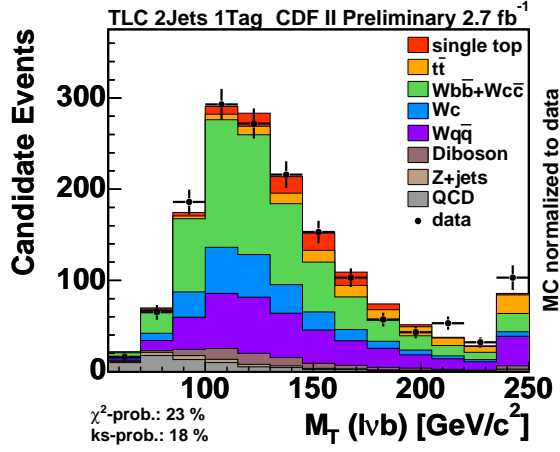


Figure 14: Data–MC comparison of the input variables four to six of the t -channel network in the 2-jet bin with 1 b -tag with 1 b -tag for TLC (left) and EMC (right) leptons.

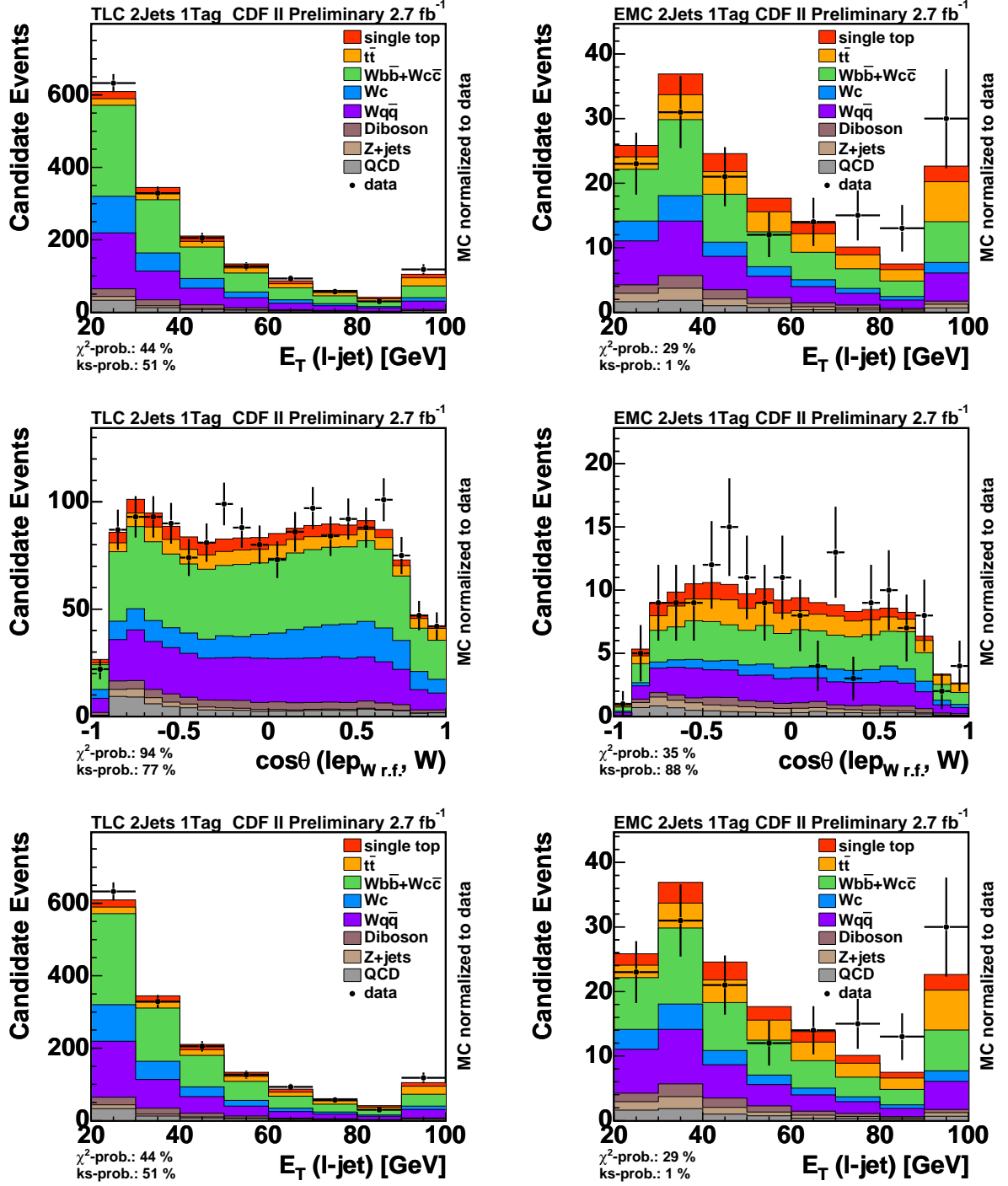


Figure 15: Data–MC comparison of the input variables seven of nine of the t -channel network in the 2-jet bin with 1 b -tag with 1 b -tag for TLC (left) and EMC (right) leptons.

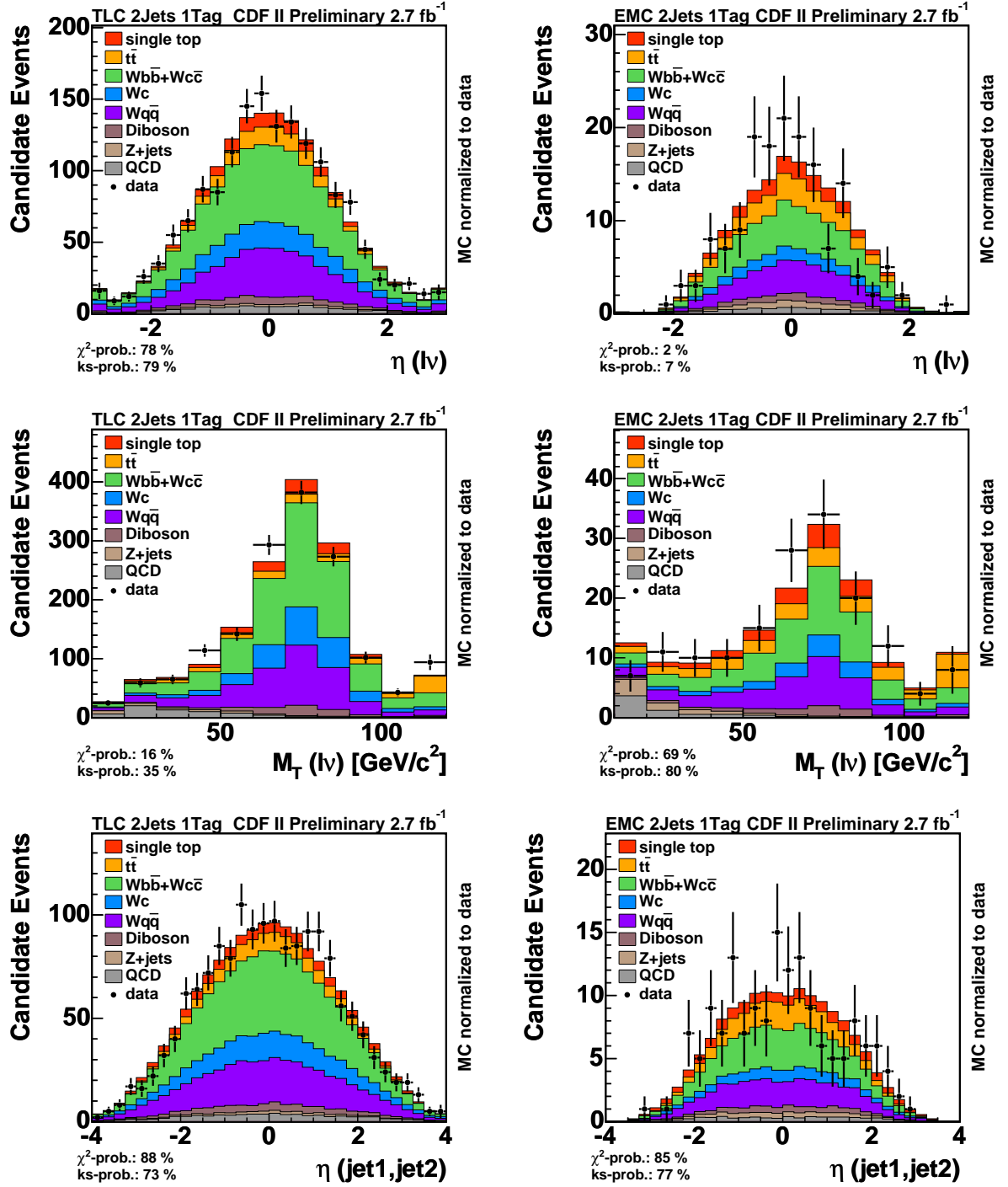


Figure 16: Data–MC comparison of the input variables ten to twelve of the t -channel network in the 2-jet bin with 1 b -tag with 1 b -tag for TLC (left) and EMC (right) leptons.

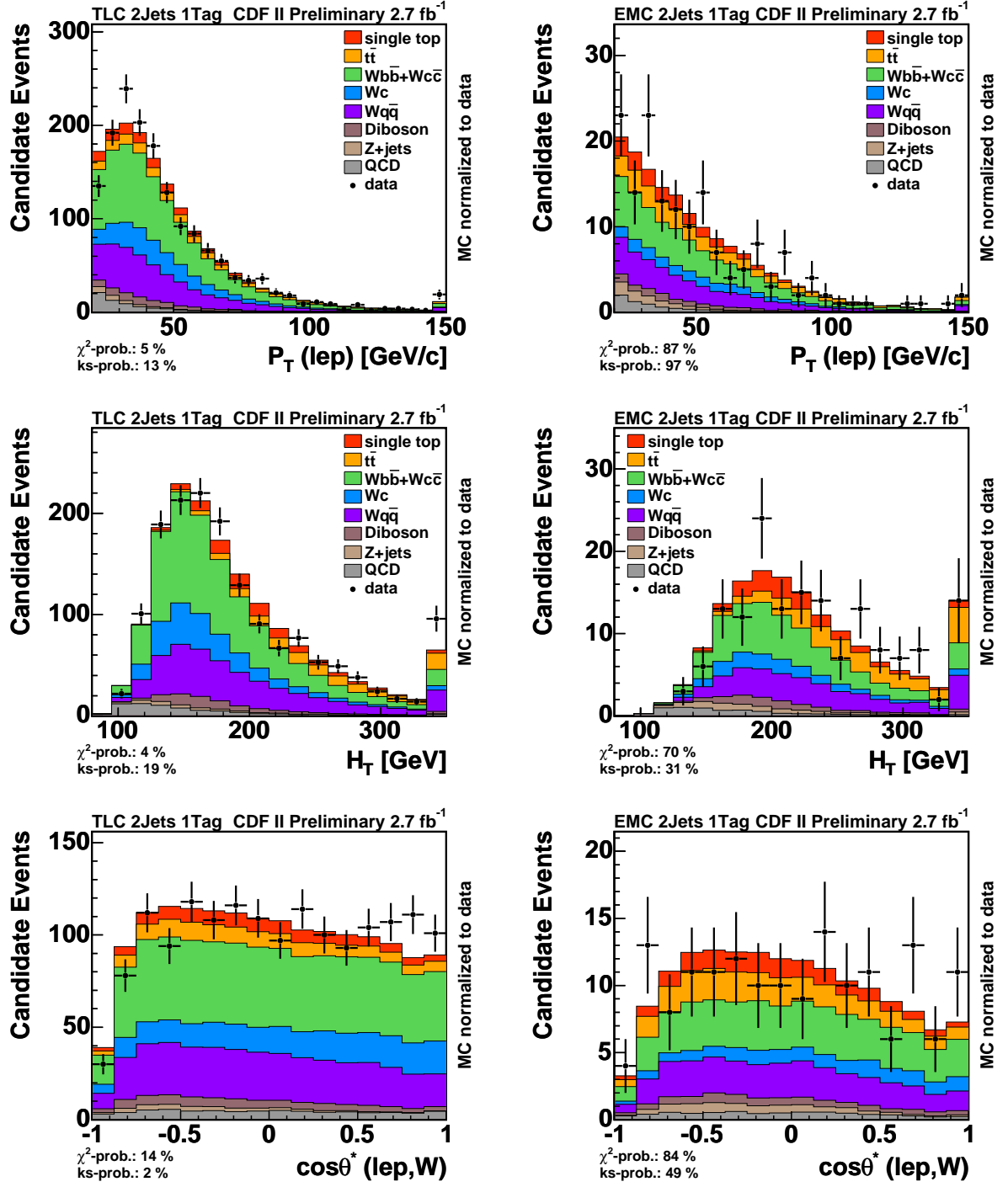


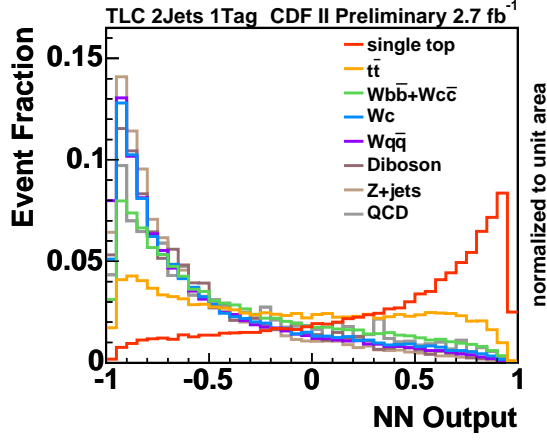
Figure 17: Data–MC comparison of the input variables thirteen to fifteen of the t -channel network in the 2-jet bin with 1 b -tag with 1 b -tag for TLC (left) and EMC (right) leptons.

5.3 Output Distributions and Templates

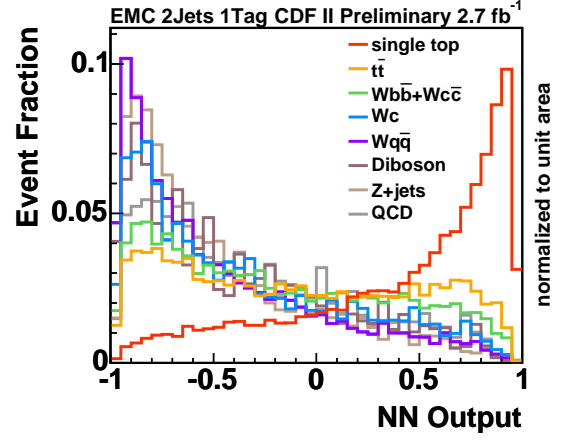
The training of a neural network results in one output variable continuously distributed between -1 and 1 . The output of the different neural networks is used to create templates which are to be fitted to the output distribution of observed events.

5.3.1 Combined Search

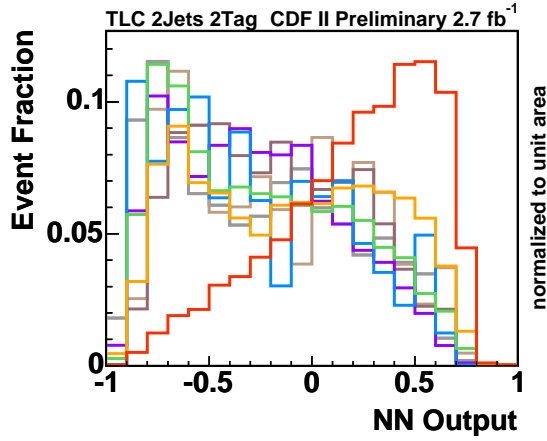
For the combined search of t - and s -channel, the following networks are utilized: the output of the networks trained to identify t -channel events are utilized in the 2-jet channel with 1 b -tag and 3-jet channel with 1 and 2 b -tags, while the output of the s -channel network is used in the 2-jet channel with 2 b -tags (figure 18 and 19). The output distributions of both t - and s -channel events are combined into one single distribution, where the ratio between the two processes is as predicted by the standard model. In the fit, all considered channels are fitted simultaneously to determine the combined single-top cross section.



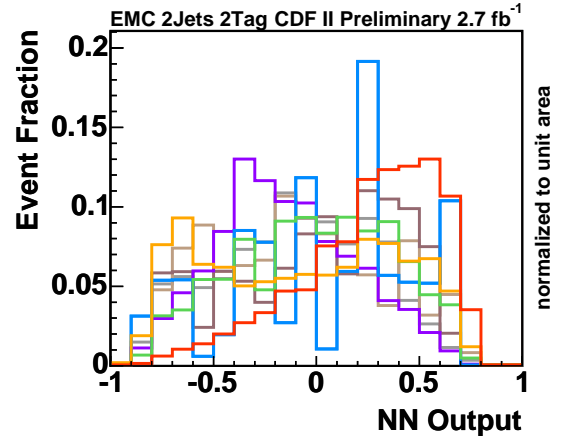
(a)



(b)

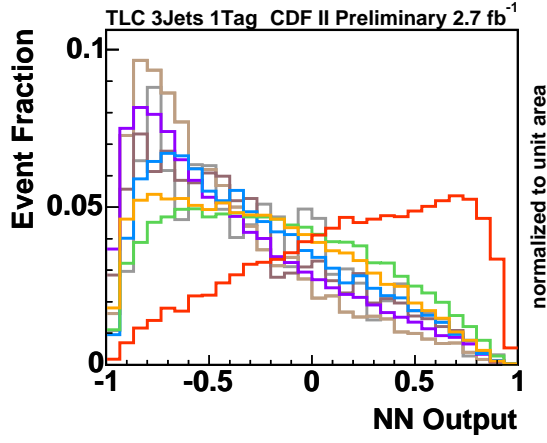


(c)

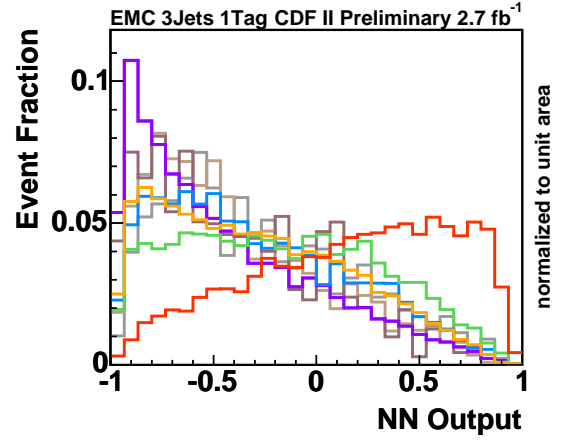


(d)

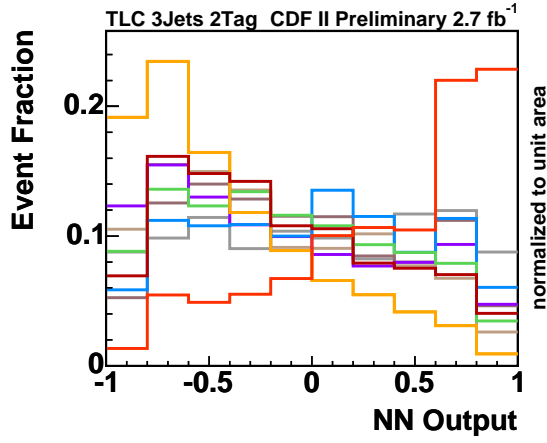
Figure 18: The templates for the t -channel neural network in the 2-jet channel with 1 b -tag (top) and the s -channel neural network in the 2-jet channel with 2 b -tags (bottom) for TLC (left) and EMC (right) leptons. The output of t - and s -channel events are added with a ratio corresponding to the standard-model prediction.



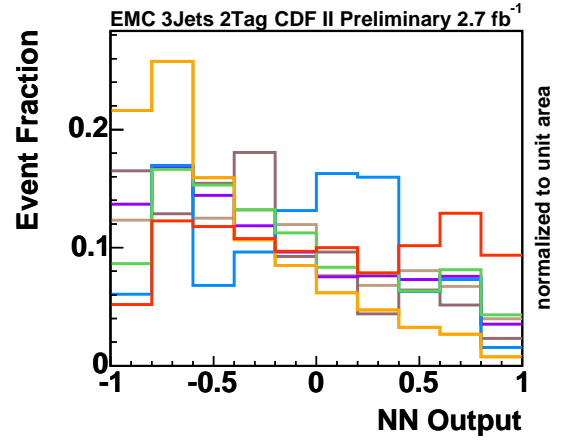
(a)



(b)



(c)



(d)

Figure 19: The templates for the t -channel neural network in the 3-jet channel with 1 b -tag (top) and with 2 b -tags (bottom) for TLC (left) and EMC (right) leptons. The output of t - and s -channel events are added with a ratio corresponding to the standard-model prediction.

5.3.2 Separate Search

The templates of the separate search in the 2-jet bin with 1 b -tag are illustrated in figures 20–23 showing the output of the s -channel neural network versus the output of the t -channel neural network. For the final fit to data the 2D templates get unwinded bin by bin to have 1D distributions. The final templates of the separate search in the 2-jet bin with 1 and 2 b -tags and 3-jet bin with 1 and 2 b -tags, respectively are illustrated in figure 24 and 25. In the fit, all considered bins are fitted simultaneously to determine the t - and s -channel cross sections.

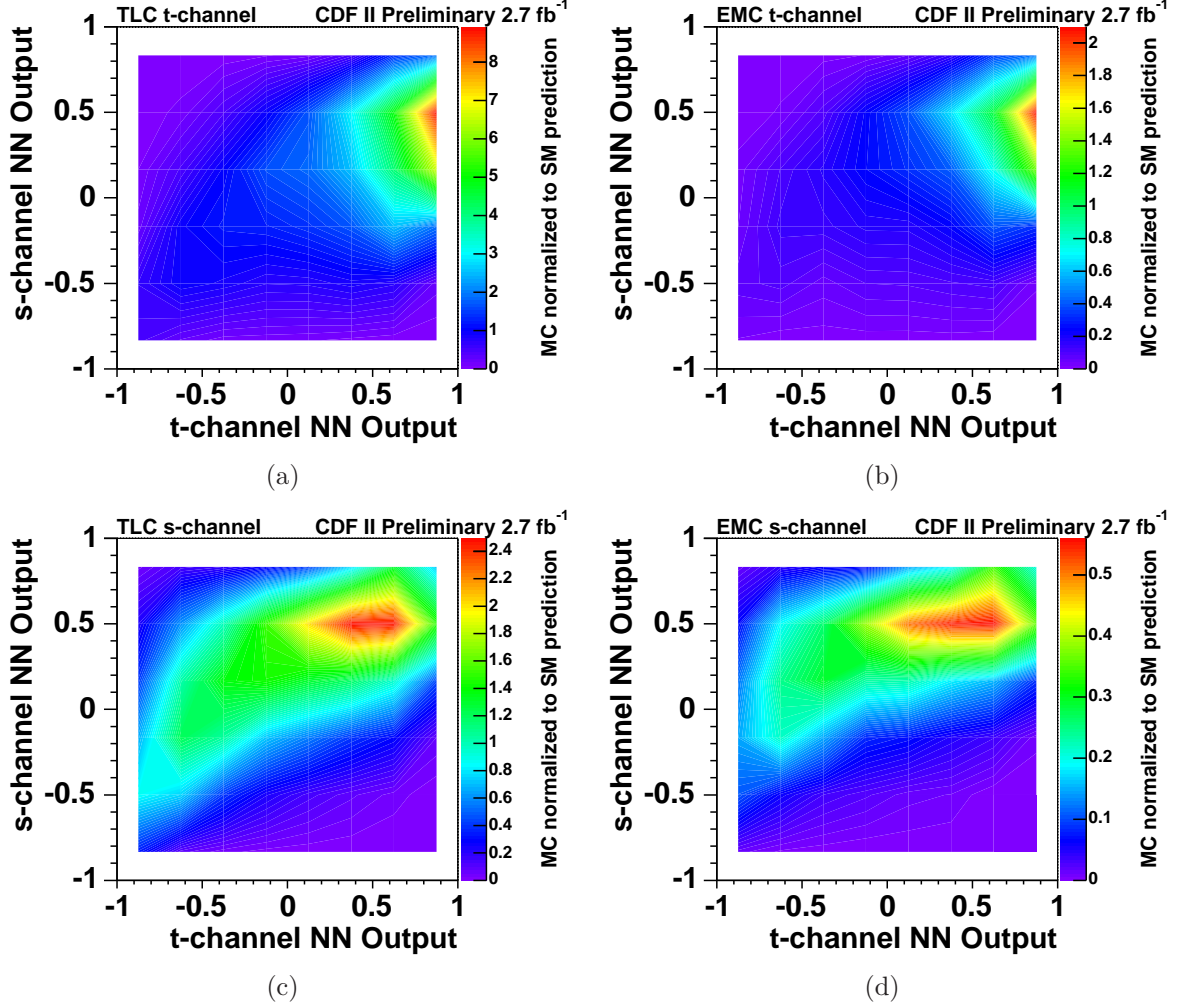


Figure 20: The templates utilized in the separate search are illustrated. The output of the s -channel network in dependence of the output of the t -channel network is shown for t -channel and s -channel events for TLC (left) and EMC (right) leptons.

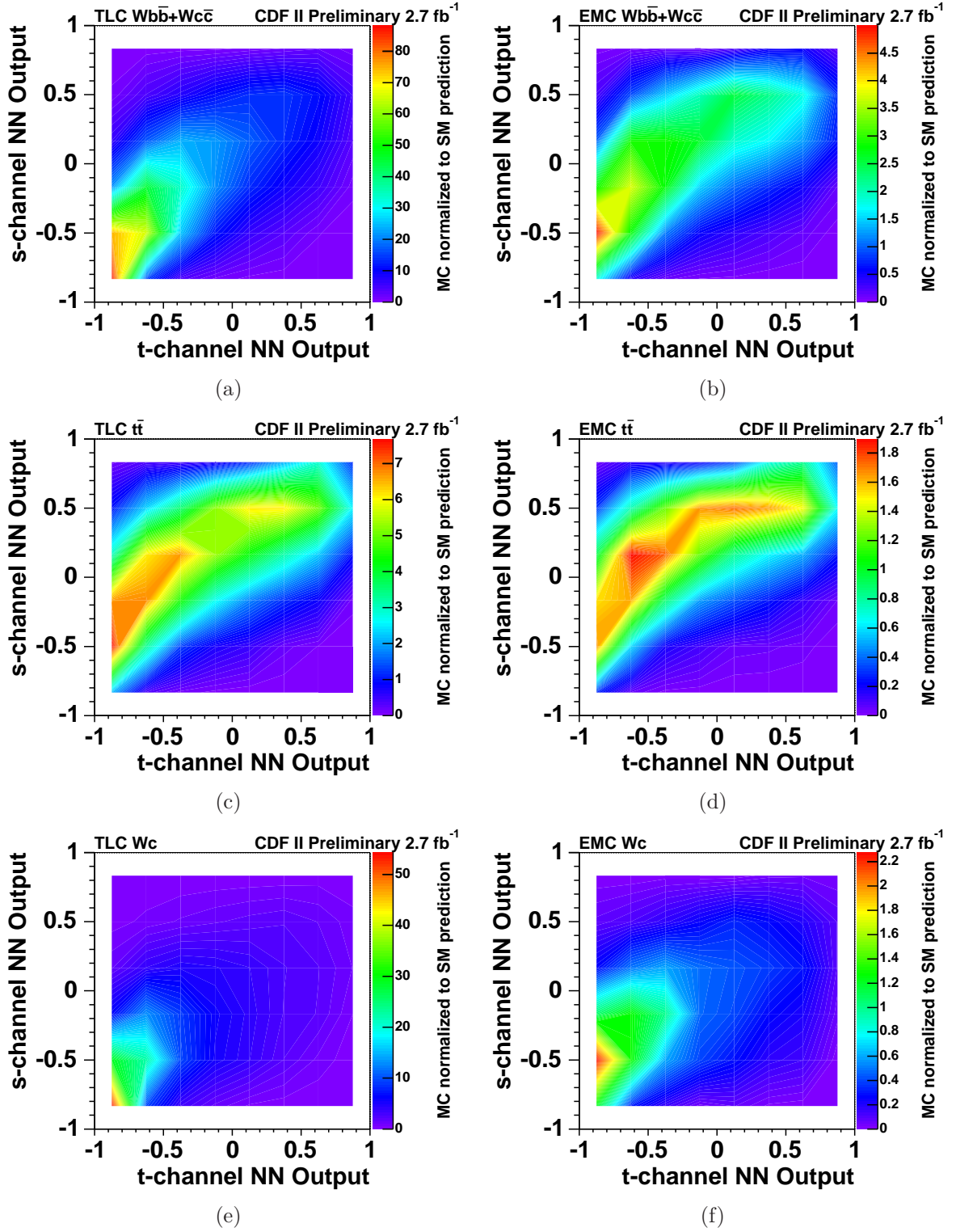


Figure 21: The templates utilized in the separate search are illustrated. The output of the s -channel network in dependence of the output of the t -channel network is shown for $Wb\bar{b} + Wc\bar{c}$ events, $t\bar{t}$ events, and Wc events for TLC (left) and EMC (right) leptons.

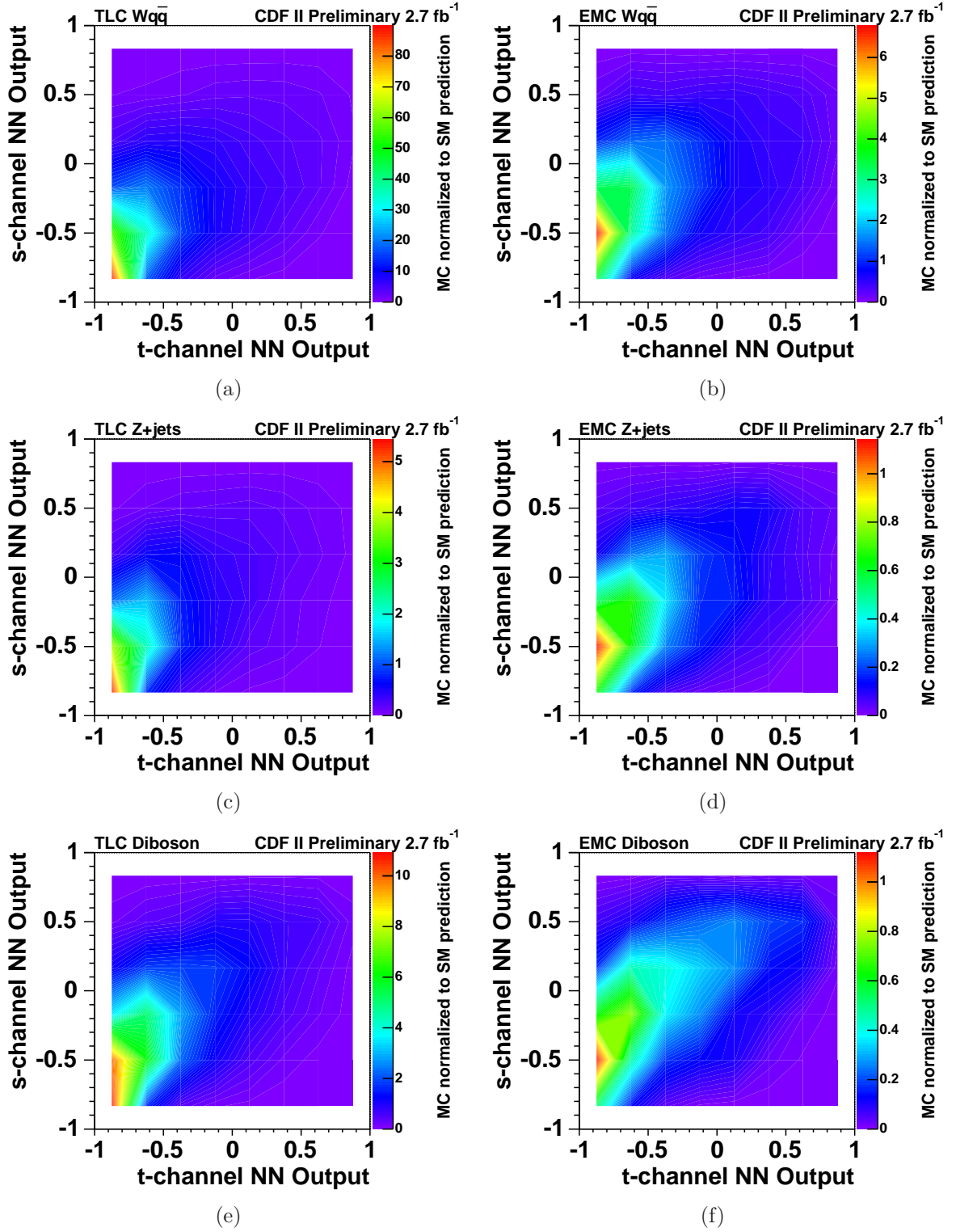


Figure 22: The templates utilized in the separate search are illustrated. The output of the s -channel network in dependence of the output of the t -channel network is shown for $Wq\bar{q}$ events, Z +jets events, and Diboson events for TLC (left) and EMC (right) leptons.

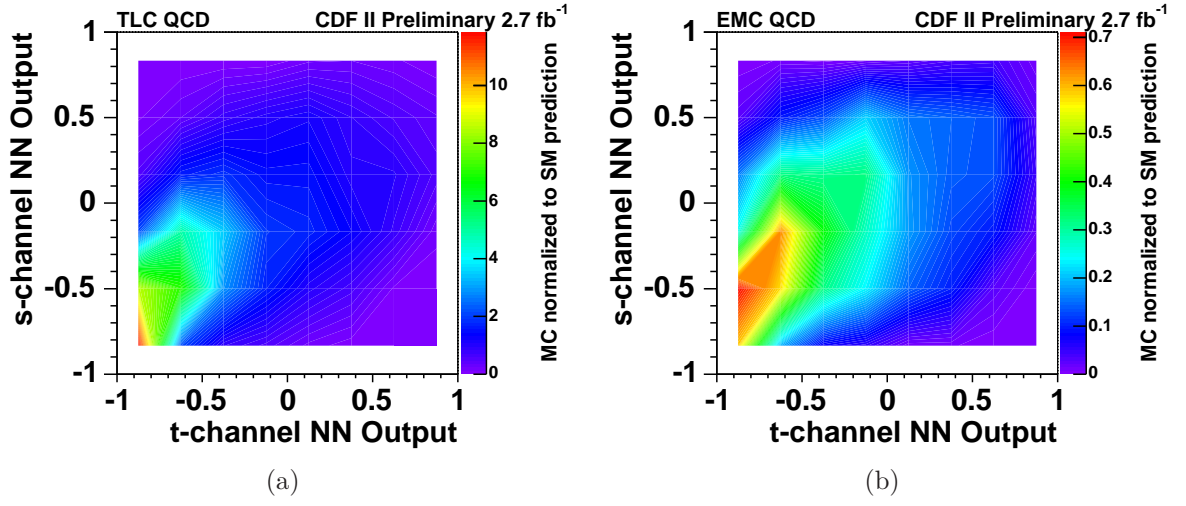


Figure 23: The templates utilized in the separate search are illustrated. The output of the s -channel network in dependence of the output of the t -channel network is shown for QCD multijet events for TLC (left) and EMC (right) leptons.

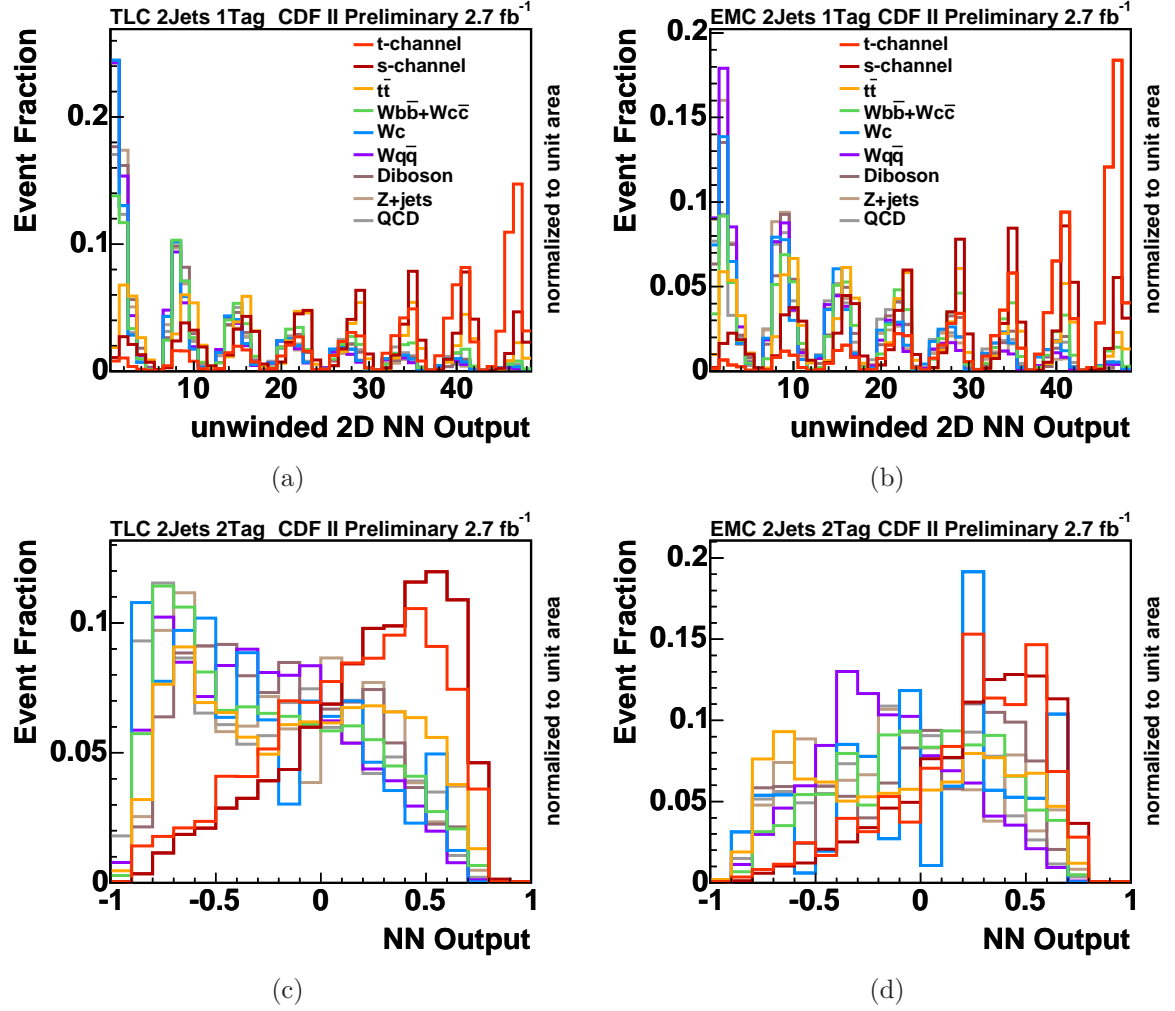


Figure 24: The templates for the separate search: the unwinded 2D NN output of the s - and t -channel neural networks in the 2-jet bin with 1 b -tag (top) and the NN output of the s -channel neural network in the 2-jet bin with 2 b -tags (bottom) for TLC (left) and EMC (right) leptons.

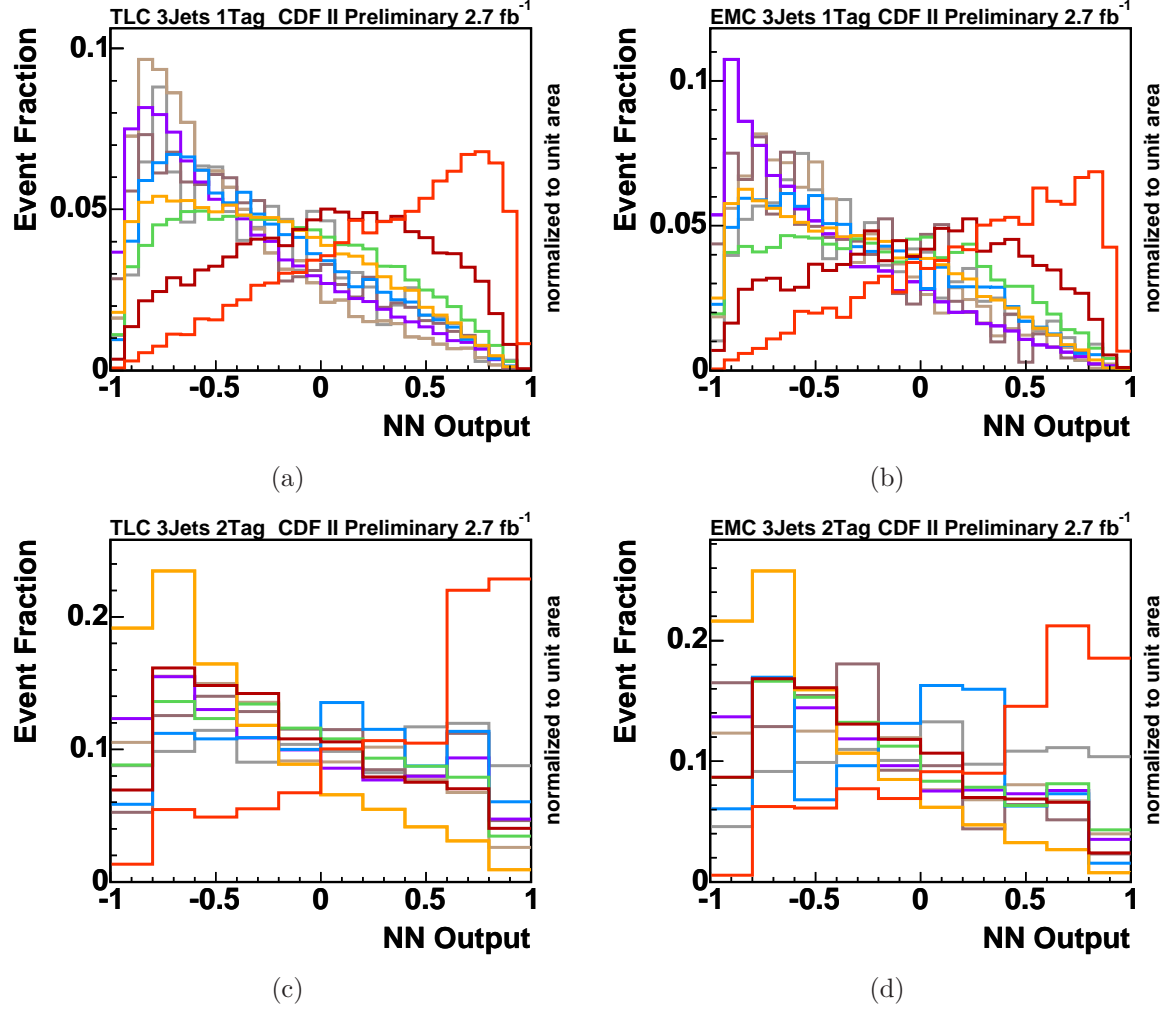


Figure 25: The templates for the separate search: the NN output of the t -channel neural networks in the 3-jet bin with 1 b -tag (top) and the 3-jet bin with 2 b -tags (bottom) for TLC (left) and EMC (right) leptons.

6 Systematic Uncertainties

Uncertainties in the modeling of physics processes and detector effects cause systematic uncertainties on the measurement results, affecting the rate of predicted signal and background events as well as the shape of the template histograms used in the fit to the observed data distribution. It must be noted that some effects induce only rate uncertainties, while some affect only the shape of the templates, even though most of them impact both.

The following sources of systematic uncertainties are considered: the uncertainty on the jet energy corrections, the uncertainty in modeling initial-state gluon radiation (ISR) and final-state gluon radiation (FSR), the choice of the parameterization of the parton distribution functions (PDF) used for the event simulation, the choice of the Monte Carlo event generator, the uncertainty in the event detection efficiency, the uncertainty in modeling the output of the neural-network b -tagger, the uncertainty in the factorization and renormalization scale for the simulation of W +heavy flavor processes, the modeling of instrumental backgrounds, that is mistag events and non- W events, the uncertainty in the luminosity determination, and the uncertainty on the modeling of the distributions of $\Delta R_{j_1, j_2}$, η_{j_2} in the 2 jet bin and η_{j_3} in the 3 jet bin in the pretag sample.

The impact of these sources of uncertainties is evaluated by altering the modeling of the corresponding processes or effects within their uncertainties or by assigning a plausible alternative model. As a result, relative changes of the event rates and shifted template distributions are obtained. Rate uncertainties are only determined for single-top-quark and $t\bar{t}$ events because the rates of the main backgrounds, W +jets and QCD events, are estimated based on the observed rate of events before b tagging or by a fit to the observed missing transverse energy distribution, respectively.

For later usage in the template likelihood fit to the observed output distribution, all shape uncertainties are smoothed using a 5 bin median filter. This means the content n_i of bin i is given by: $median(n_{i-2}, n_{i-1}, n_i, n_{i+1}, n_{i+2})$.

As one can see from the formula this can only be done for all bins, but not for the first two and last two bins. The median filter has the advantages to remove extrem outliers and high frequency noise very efficiently, but doesn't change already smooth distributions. In figure 26 a comparison between the original and the smoothed ratio plots is shown.

The effect of the uncertainty on the jet energy scale (JES) corrections is quantified by varying the corrections within their $\pm 1\sigma$ uncertainties [?]. The corresponding alternative template distributions are calculated for all signal and background processes and are shown in figures 27 and 28.

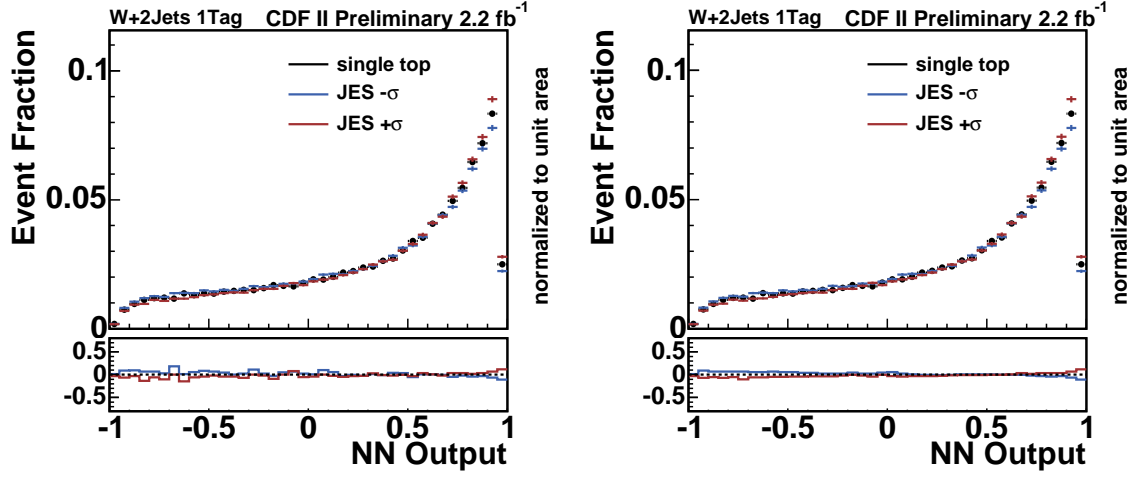


Figure 26: Comparison of the original (left) and smoothed (right) relative difference between the shifted distribution and the default distribution.

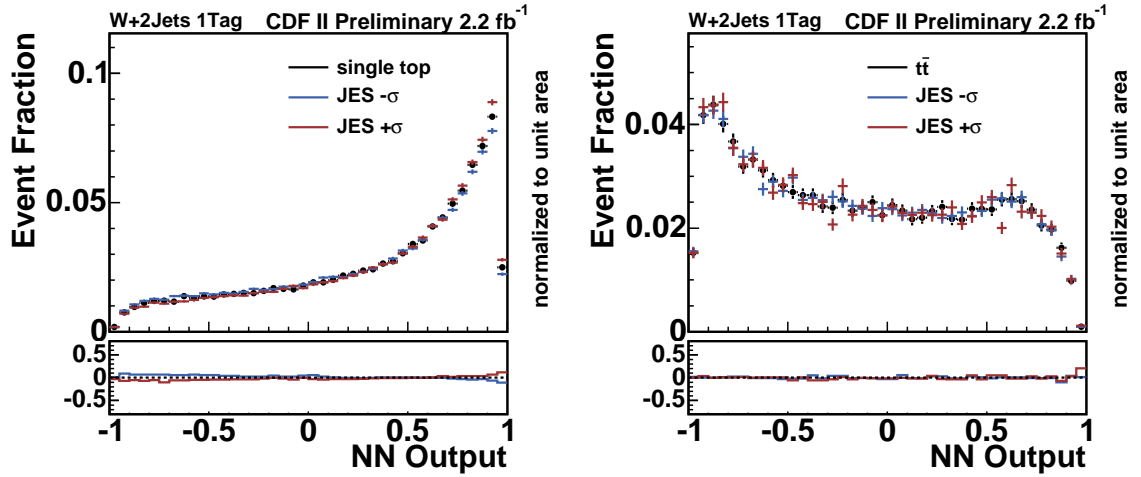


Figure 27: Shape systematics due to the uncertainty on the jet energy scale (JES) correction for single-top and $t\bar{t}$. In the upper plot the default distribution is shown in comparison to the shifted distributions. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

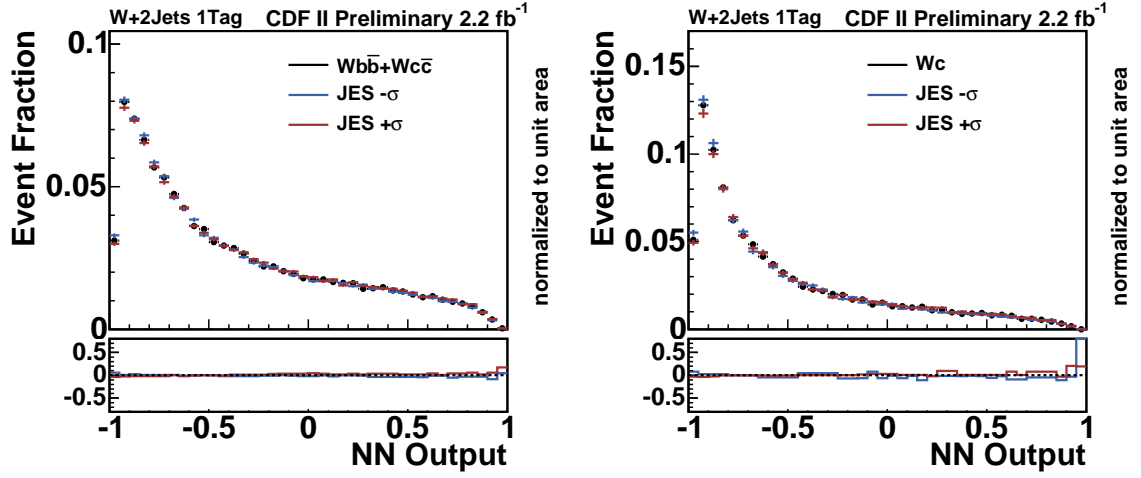


Figure 28: Shape systematics due the uncertainty on the jet energy scale (JES) correction for $WbbWcc$ and Wc . In the upper plot the default distribution is shown in comparison to the shifted distributions. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

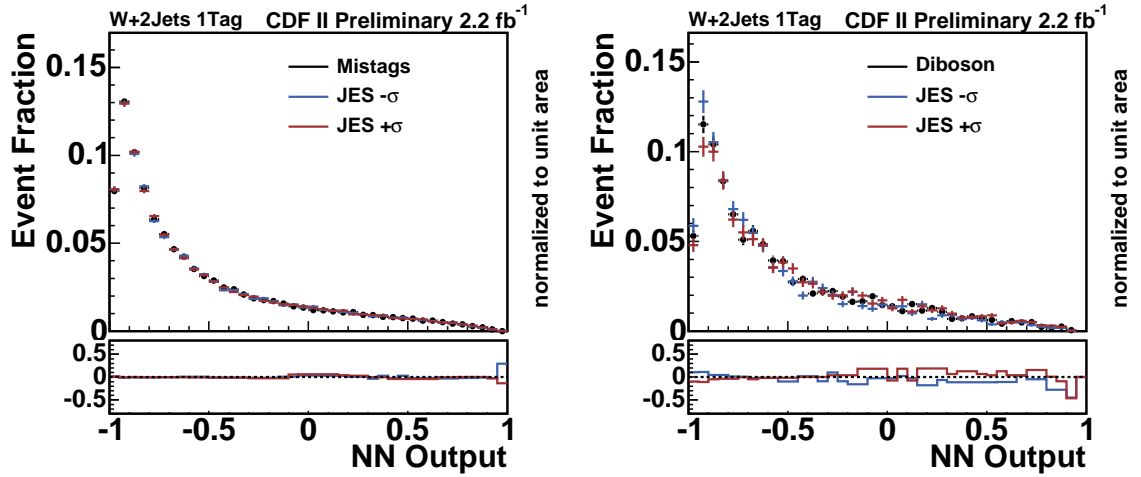


Figure 29: Shape systematics due the uncertainty on the jet energy scale (JES) correction for mistags and diboson. In the upper plot the default distribution is shown in comparison to the shifted distributions. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

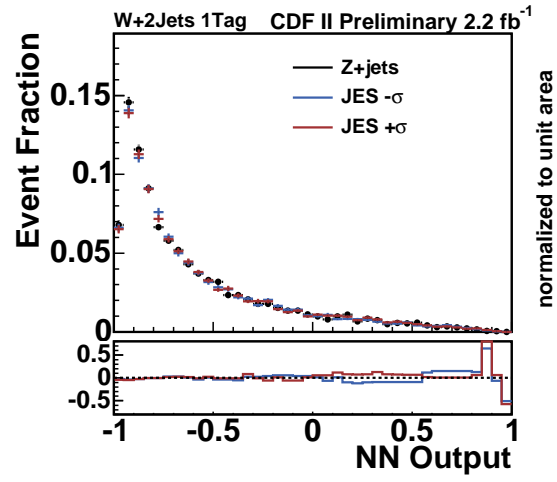


Figure 30: Shape systematics due to the uncertainty on the jet energy scale (JES) correction for Z +jets. In the upper plot the default distribution is shown in comparison to the shifted distributions. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

The influence of initial-state and final-state gluon radiation is estimated by producing samples of simulated events for which the simulation was altered to produce either less or more gluon radiation compared to the standard setting [?]. Specifically, two parameters controlling the parton shower in the PYTHIA program are varied: Λ_{QCD} and the scale factor K to the transverse momentum scale of the showering. The different settings are derived from studies of ISR in Drell-Yan events. Using these specific ISR and FSR samples of simulated events, alternative template shapes are produced for single-top-quark and $t\bar{t}$ events as illustrated in figures 31 and 32.

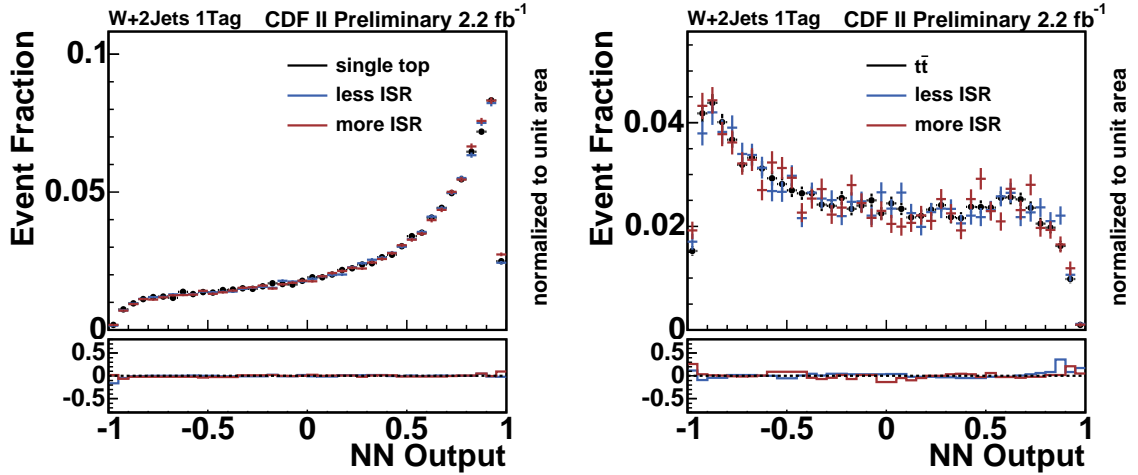


Figure 31: Shape systematics due to the uncertainty in the final state radiation for single-top and $t\bar{t}$. In the upper plot the default distribution is shown in comparison to the shifted distributions. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

The impact of the uncertainties on the PDF parameterization are studied by reweighting single-top-quark and $t\bar{t}$ events with weights associated with the 20 pairs of CTEQ6M eigenvectors. The rate uncertainty on the signal model which is based on the MADEVENT event generator is determined by a comparison to differential cross sections computed with the ZTOP program. The modeling of $t\bar{t}$ events is studied by using simulated events produced with the MC@NLO program and showered by HERWIG as an alternative generator. The factorization and renormalization scale is varied in the simulation to derive an additional set of altered template histograms for W +heavy flavor events. The default W +jets Monte Carlo samples are generated with a dynamic scale $\mu^2 = Q^2$. The influence of the variation of Q^2 on the shape of the output distribution of $Wb\bar{b}$ events is illustrated in figure 33.

The uncertainty in the event detection efficiency ϵ_{evt} includes the uncertainties on the trigger efficiency, on the lepton identification efficiency, and on the b -tagging efficiency which is the dominating factor. Since no cut is applied on the output of the neural-network b -tagger, the uncertainty associated with this quantity does not imply a rate uncertainty, but only a shape uncertainty on the template distributions. Systematic effects are studied by utilizing the correction function derived for the mistags. Therefore we consider two

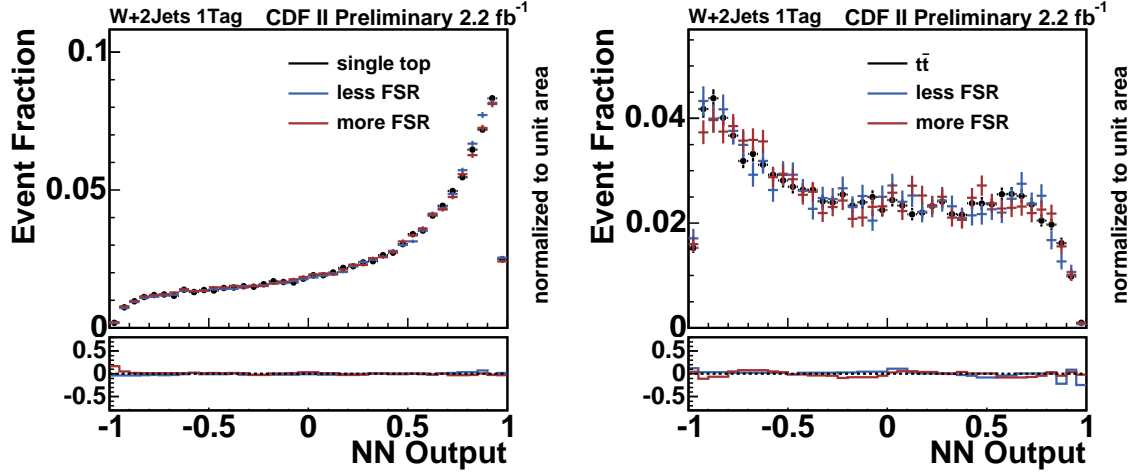


Figure 32: Shape systematics due to the uncertainty in the final state radiation for single-top and $t\bar{t}$. In the upper plot the default distribution is shown in comparison to the shifted distributions. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

scenarios. The pessimistic one, in which we apply the correction function on the c-like templates, that they get more signal like. And the optimistic one, in which we use the uncorrected mistag shape, that the mistag template gets more background like.

The influence of these scenarios on the mistags and Wc templates is illustrated in figure 34. A modified model of QCD events is considered to investigate the influence of this aspect on the analysis. In comparison to the default model described in section 4, the alternative model uses events selected from a generic jet sample where one of the jets has a fraction of electromagnetic energy measured in the calorimeter of at least 0.8, but less than 0.95. This so-called jet-electron assumes the role of the charged lepton in the event. In addition, the flavor composition of the QCD sample is varied: the default model assumes a composition of 45% b -quark jets, 40% c -quark jets, and 15% light-quark jets, whereas the alternative model uses a composition of 60:30:10 [?], respectively. The influence of the flavor composition on the nonW template is illustrated in figure 35. To evaluate the systematic effect on the shapes of the distributions caused by the modeling of mistagged light-quark jet events, an alternative model is utilized to create template distributions. This is realized by replacing the default mistag model based on simulated events as described in section 4 by a description on the basis of measured W +jets events before b tagging. The influence of the mistag model is illustrated in figure 36. The uncertainty on the modeling of the distributions of ΔR and η_{j2} in the pretag sample is determined by reweighting the distributions. The influence of these reweightings are in figures 37, 38, 39 and 40 shown. Tables 9-17 summarize the relative rate uncertainties on the event prediction of the various Monte Carlo samples.

The analyses are done under the assumption of a top-quark mass of $M_t = 175 \text{ GeV}/c^2$. That is why the uncertainty in the top-quark mass is not taken into account as a systematic uncertainty. Hence, the analyses provide rather a measurement at the specified value

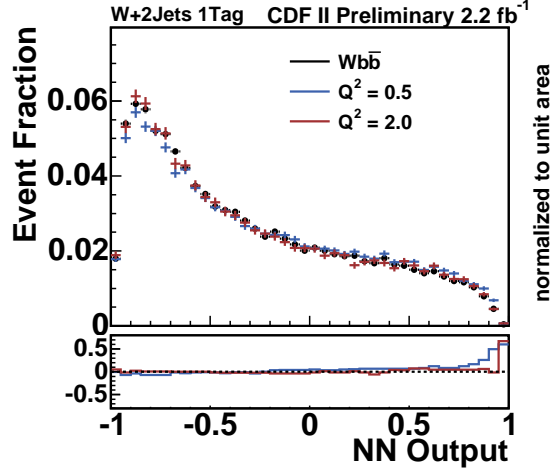


Figure 33: Shape systematics due to the uncertainty in the ALPGEN factorization/renormalization scale Q^2 for the $Wb\bar{b}$ background. The upper plot shows the default distribution in comparison to the shifted distribution. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

of the top-quark mass. However, if the top-quark mass is varied in the simulation by $\pm 5 \text{ GeV}/c^2$, the acceptance for single-top-quark events changes as shown in table 9-16.

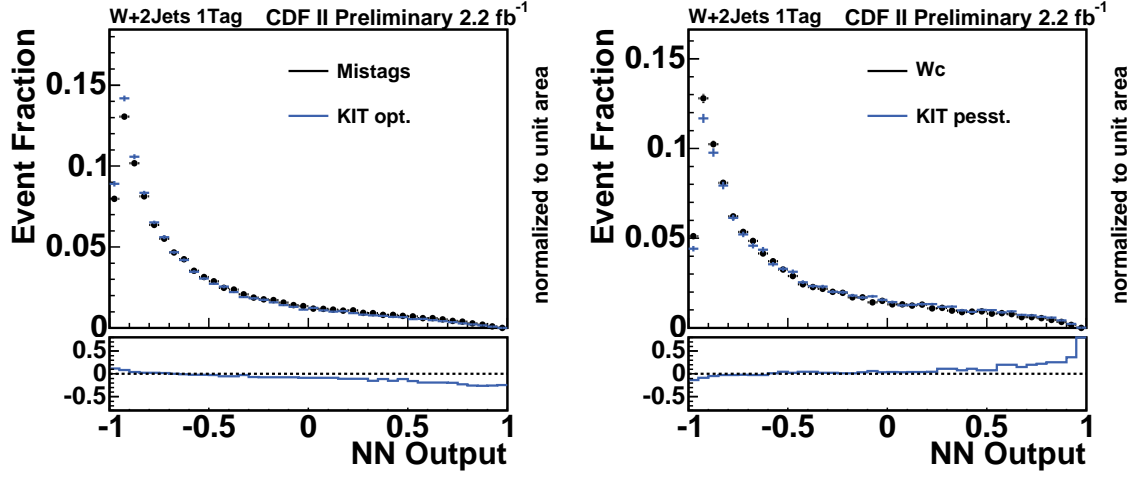


Figure 34: Shape systematics due to the influence of the mistags correction function of the KIT flavor separator on mistags and Wc . The upper plot shows the default distribution in comparison to the shifted distribution. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

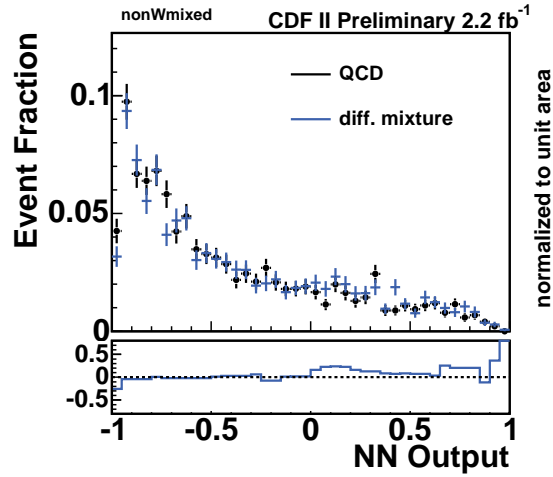


Figure 35: Shape systematics due to the influence of the nonW flavor composition. The upper plot shows the default distribution in comparison to the shifted distribution. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

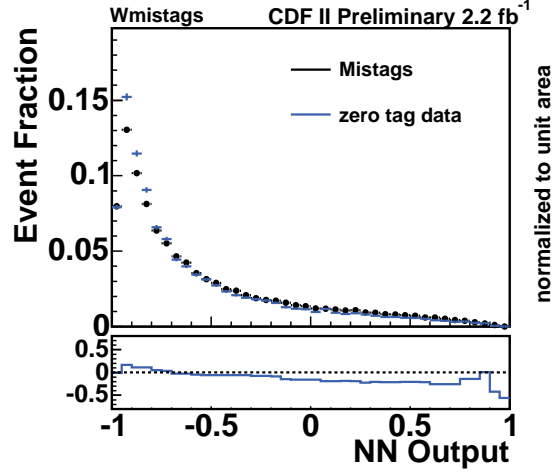


Figure 36: Shape systematics due to the influence of the mistag model. The upper plot shows the default distribution in comparison to the shifted distribution. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

Source	t -channel	s -channel	single-top	$t\bar{t}$
IFSR less/more	7.0/-1.5 %	6.2/7.1 %	6.7/1.4 %	-7.7/-9.7 %
PDF	3.1/-3.5 %	1.7/-1.4 %	2.6/-2.8 %	1.9/-2.3 %
MC	2.0/-2.0 %	1.0/-1.0 %	1.7/-1.7 %	-2.7/2.7 %
ϵ_{evt}	4.2/-4.2 %	2.3/-2.3 %	3.6/-3.6 %	2.9/-2.9 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.6/-12.6 %	12.4/-12.4 %
M_{top} 170/180	6.1/-5.3 %	9.5/-8.0 %	7.3/-6.2 %	7.8/-8.1 %
	Diboson	Z+jets		
ϵ_{evt}	7.6/-7.6 %	8.3/-8.3 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
Cross section	1.9/-1.9 %	10.8/-10.8 %		

Table 9: Systematic rate uncertainties for TLC lepton events with 2 jets and 1 b tag

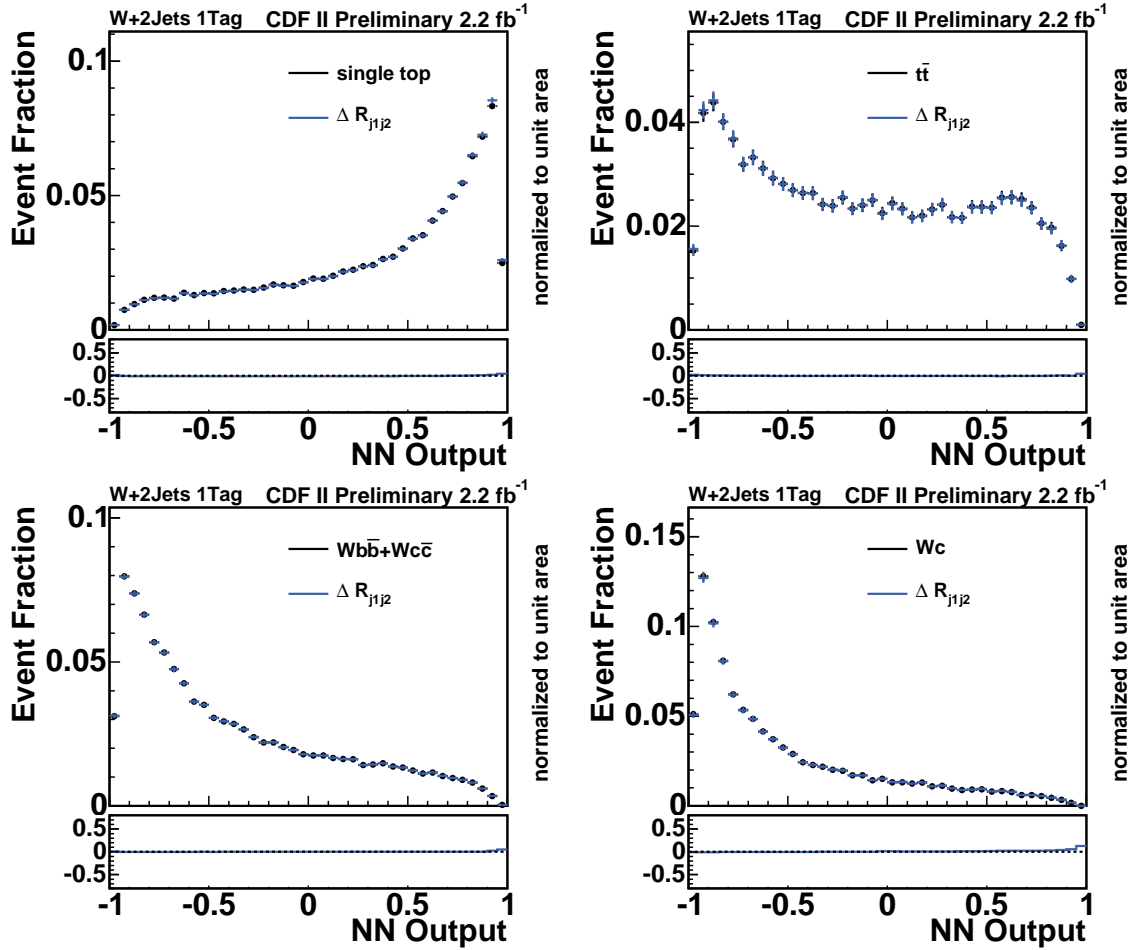


Figure 37: Shape systematics due to the influence of the reweightings due to mismodeling in ΔR between the first and the second jet for the single top, $t\bar{t}$, $WbbWcc$ and Wc templates. The upper plots show the default distribution in comparison to the shifted distribution. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

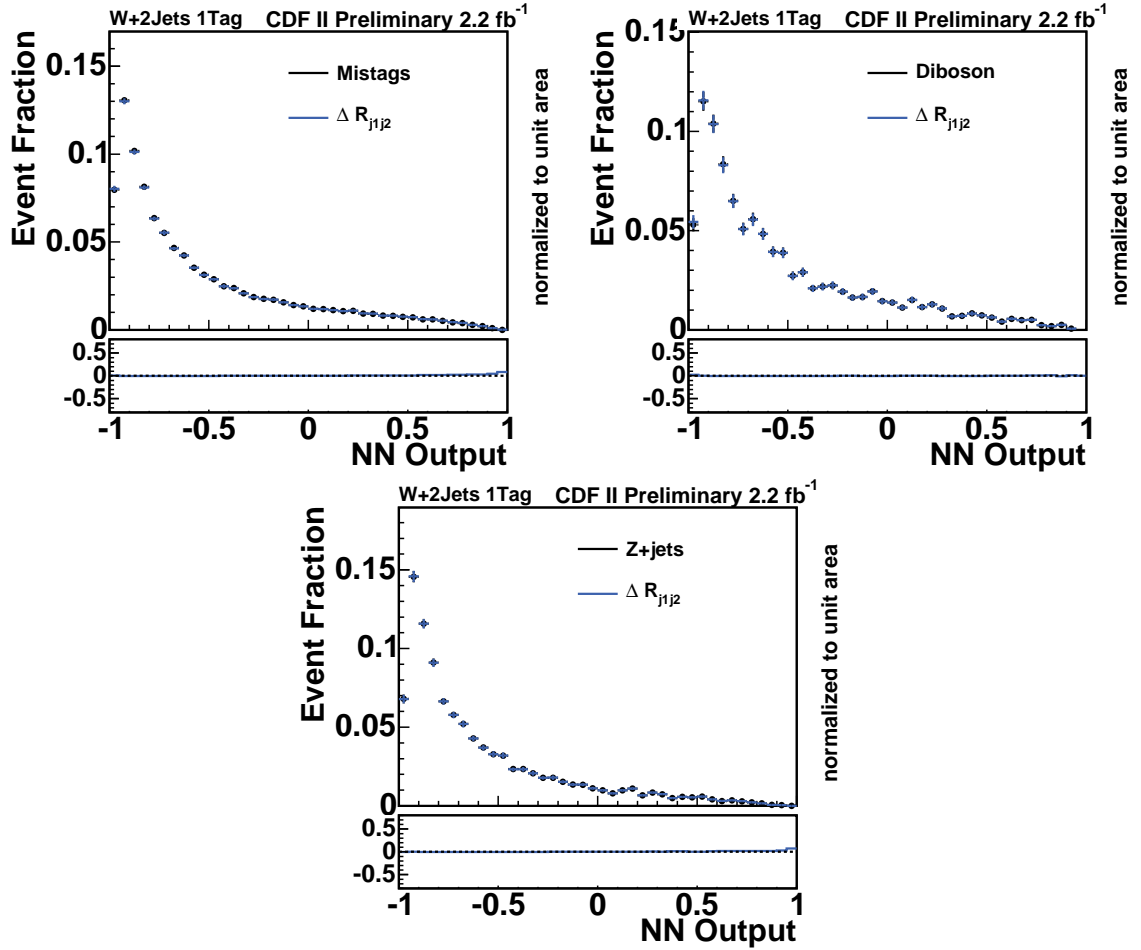


Figure 38: Shape systematics due the influence of the reweightings due to mismodeling in ΔR between the first and the second jet for the mistag, diboson and Z +jets templates. The upper plots show the default distribution in comparison to the shifted distribution. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

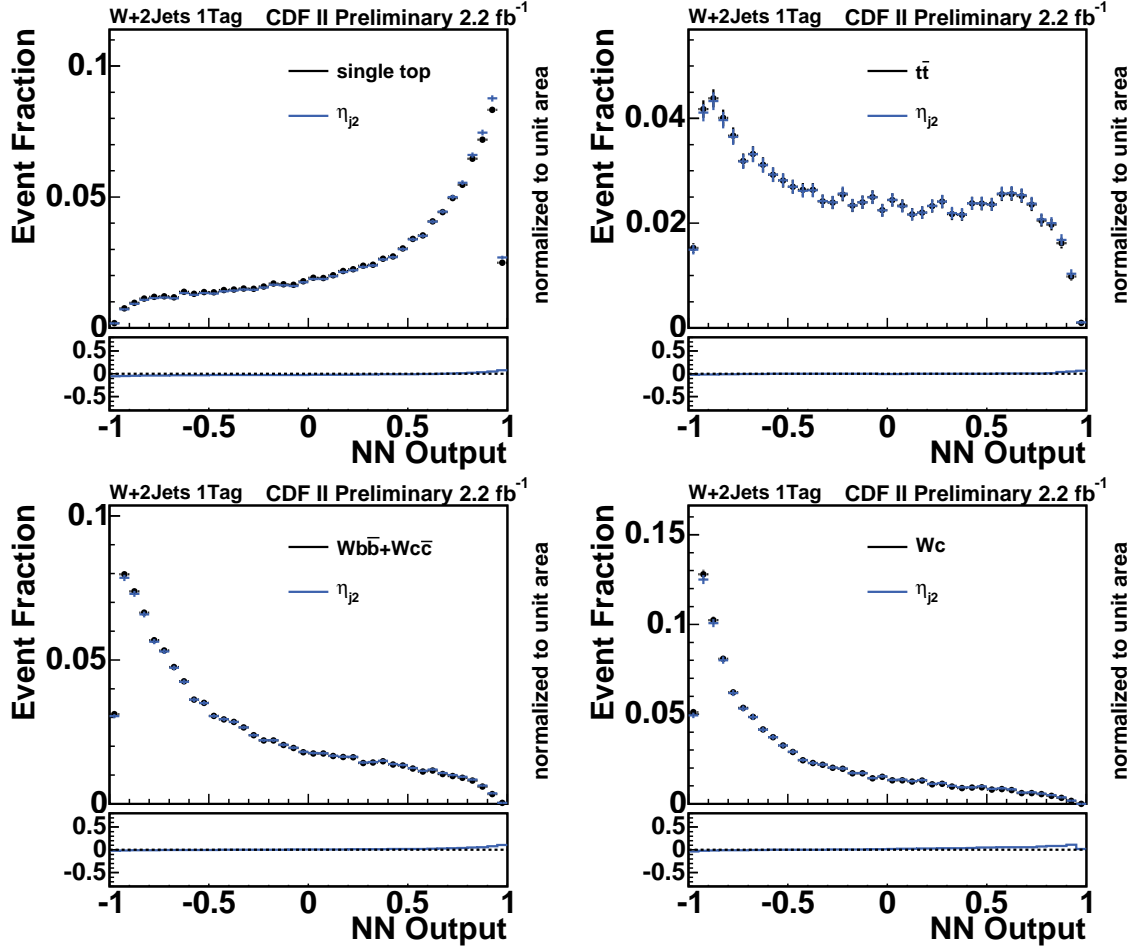


Figure 39: Shape systematics due to the influence of the reweightings due to mismodeling of the η of the second jet for the single top, $t\bar{t}$, $Wb\bar{b}Wc\bar{c}$ and Wc templates. The upper plots show the default distribution in comparison to the shifted distribution. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

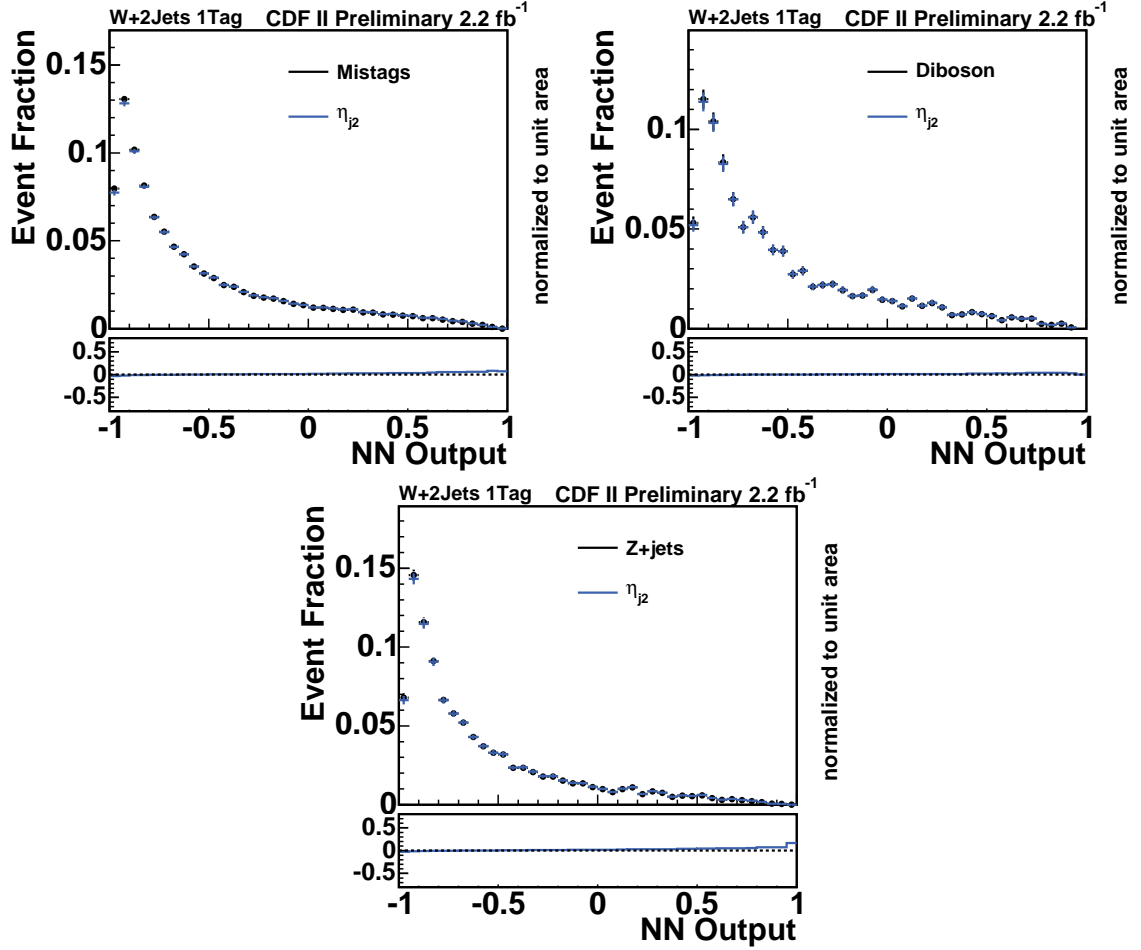


Figure 40: Shape systematics due the influence of the reweightings due to mismodeling of the η of the second jet between the first and the second jet for the mistag, diboson and Z +jets templates. The upper plots show the default distribution in comparison to the shifted distribution. In the lower plot the relative difference between the shifted distribution and the default is plotted, which is smoothed using a 5 bin median smoothing procedure.

Source	t -channel	s -channel	single-top	$t\bar{t}$
IFSR less/more	7.0/-1.5 %	6.2/7.1 %	6.7/1.6 %	-7.7/-9.7 %
PDF	3.1/-3.5 %	1.7/-1.4 %	2.6/-2.7 %	1.9/-2.3 %
MC	2.0/-2.0 %	1.0/-1.0 %	1.6/-1.6 %	2.7/-2.7 %
ϵ_{evt}	3.9/-3.9 %	1.5/-1.5 %	3.0/-3.0 %	2.6/-2.6 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.5/-12.5 %	12.4/-12.4 %
M_{top} 170/180	6.1/-5.3 %	9.5/-8.0 %	7.3/-6.3 %	7.8/-8.1 %
	Diboson	Z+jets		
ϵ_{evt}	7.2/-7.2 %	7.6/-7.6 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
Cross section	1.9/-1.9 %	10.8/-10.8 %		

Table 10: Systematic rate uncertainties for EMC lepton events with 2 Jets and 1 b tags

Source	t -channel	s -channel	single-top	$t\bar{t}$
IFSR less/more	-0.9/-13.4 %	9.5/11.4 %	7.9/7.5 %	-7.5/-11.2 %
PDF	3.0/-3.3 %	1.7/-1.5 %	1.9/-1.7 %	1.9/-2.3 %
MC	2.0/-2.0 %	1.0/-1.0 %	1.2/-1.2 %	4.6/-4.6 %
ϵ_{evt}	10.0/-10.0 %	8.7/-8.7 %	8.9/-8.9 %	9.0/-9.0 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.5/-12.5 %	12.4/-12.4 %
M_{top} 170/180	2.5/-7.3 %	9.4/-6.9 %	7.7/-6.7 %	9.9/-7.1 %
	Diboson	Z+jets	Mistags	
ϵ_{evt}	9.8/-9.8 %	10.6/-10.6 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
Double tag			22.0/-22.0%	
Cross section	1.9/-1.9 %	10.8/-10.8 %		

Table 11: Systematic rate uncertainties for TLC lepton events with 2 jets and 2 b tags

Source	t -channel	s -channel	single-top	$t\bar{t}$
IFSR less/more	-1.0/-13.5 %	9.4/11.4 %	8.0/8.2 %	-7.4/-11.1 %
PDF	3.0/-3.3 %	1.7/-1.5 %	1.9/-1.7 %	1.9/-2.3 %
MC	2.0/-2.0 %	1.0/-1.0 %	1.1/-1.1 %	-4.6/4.6 %
ϵ_{evt}	10.0/-10.0 %	8.7/-8.7 %	8.9/-8.9 %	8.9/-8.9 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.4/-12.4 %	12.4/-12.4 %
M_{top} 170/180	2.5/-7.3 %	9.4/-6.9 %	8.5/-7.0 %	9.9/-7.1 %
	Diboson	Z+jets	Mistags	
ϵ_{evt}	9.3/-9.3 %	9.9/-9.9 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
double tag			22.0/-22.0 %	
Cross section	3.2/-3.2 %	10.8/-10.8 %		

Table 12: Systematic rate uncertainties for EMC lepton events with 2 Jets and 2 b tags

Source	t -channel	s -channel	single-top	$t\bar{t}$
IFSR less/more	-8.4/-3.1 %	-3.6/-17.4 %	-6.5/-8.6 %	-4.0/-6.7 %
PDF	3.2/-3.7 %	1.8/-1.5 %	2.7/-2.8 %	1.9/-2.3 %
MC	1.9/-1.9 %	1.5/-1.5 %	1.7/-1.7 %	-1.7/1.7 %
ϵ_{evt}	3.5/-3.5 %	2.3/-2.3 %	3.0/-3.0 %	2.3/-2.3 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.6/-12.6 %	12.4/-12.4 %
M_{top} 170/180	6.2/-6.5 %	11.7/-8.6 %	6.4/-5.6 %	9.3/-8.4 %
	Diboson	Z+jets		
ϵ_{evt}	7.8/-7.8 %	7.8/-7.8 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
Cross section	1.9/-1.9 %	10.8/-10.8 %		

Table 13: Systematic rate uncertainties for TLC lepton events with 3 jets and 1 b tag

Source	t -channel	s -channel	single-top	$t\bar{t}$
IFSR less/more	-8.3/-3.1 %	-3.6/-17.4 %	-6.4/-9.0 %	-4.1/-6.7 %
PDF	3.2/-3.7 %	1.7/-1.5 %	2.6/-2.8 %	1.9/-2.3 %
MC	1.5/-1.5 %	1.9/-1.9 %	1.7/-1.7 %	2.7/-2.7 %
ϵ_{evt}	3.0/-3.0 %	1.5/-1.5 %	2.4/-2.4 %	2.5/-2.5 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.5/-12.5 %	12.4/-12.4 %
M_{top} 170/180	6.2/-6.5 %	11.7/-8.6 %	8.5/-7.4 %	9.3/-8.4 %
	Diboson	Z+jets		
ϵ_{evt}	7.3/-7.3 %	14.2/-11.2 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
Cross section	1.9/-1.9 %	10.8/-10.8 %		

Table 14: Systematic rate uncertainties for EMC lepton events with 3 Jets and 1 b tags

Source	t -channel	s -channel	single-top	$t\bar{t}$
IFSR less/more	22.7/4.4 %	-3.1/-16.2 %	8.2/-7.2 %	-3.9/-9.3 %
PDF	3.7/-4.1 %	1.8/-1.5 %	2.6/-2.6 %	1.9/-2.3 %
MC	1.9/-1.9 %	1.5/-1.5 %	1.7/-1.7 %	2.0/-2.0 %
ϵ_{evt}	9.1/-9.1 %	8.8/-8.8 %	8.9/-8.9 %	9.1/-9.1 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.5/-12.5 %	12.4/-12.4 %
M_{top} 170/180	7.8/-3.0 %	9.1/-11.1 %	6.9/-6.0 %	9.3/-9.5 %
	Diboson	Z+jets	Mistags	
ϵ_{evt}	10.8/-10.8%	11.1/-11.1 %		
Luminosity	6.0/-6.0 %	6.0/-6.0 %		
Double tag			22.0/-22.0%	
Cross section	1.9/-1.9 %	10.8/-10.8 %		

Table 15: Systematic rate uncertainties for TLC lepton events with 3 jets and 2 b tags

Source	t -channel	s -channel	single-top	$t\bar{t}$
IFSR less/more	22.8/4.5 %	-3.1/-16.2 %	8.0/-7.3 %	-3.9/-9.3 %
PDF	3.7/-4.1 %	1.8/-1.5 %	2.6/-2.6 %	1.9/-2.3 %
MC	1.5/-1.5 %	1.9/-1.9 %	1.7/-1.7 %	-2.0/2.0 %
ϵ_{evt}	9.0/-9.0 %	8.8/-8.8 %	8.9/-8.9 %	9.1/-9.1 %
Luminosity	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %	6.0/-6.0 %
Cross section	12.6/-12.6 %	12.4/-12.4 %	12.5/-12.5 %	12.4/-12.4 %
M_{top} 170/180	7.8/-3.0 %	9.1/-11.1 %	8.5/-7.6 %	9.3/-9.5 %
	Diboson	Z+jets	Mistags	
ϵ_{evt}	11.5/-11.5 %	10.8/-10.8 %		
Luminosity double tag	6.0/-6.0 %	6.0/-6.0 %	22.0/-22.0 %	
Cross section	1.7/-1.7 %	10.8/-10.8 %		

Table 16: Systematic rate uncertainties for EMC lepton events with 3 Jets and 2 b tags

process	2jets 1tag	2jets 2tags	3jets 1tag	3jets 2tags
t -ch	-2.3/2.4 %	1.9/-1.7 %	-10.1/12.7 %	-3.9/5.8 %
s -ch	-1.7/0.4 %	-1.3/0.4 %	-9.8/8.6 %	-6.9/10.4 %
single-top	-2.1/1.7 %	-0.9/0.1 %	-10.0/11.0 %	-5.6/8.4 %
$t\bar{t}$	9.1/-8.9 %	6.1/-6.4 %	3.8/-4.2 %	4.1/-4.5 %
$Wc\bar{c}+Wb\bar{b}$	5.5/-5.4 %	7.0/-6.3 %	6.4/-6.4 %	10.4/-7.1 %
Wc	5.7/-4.9 %	6.9/-4.6 %	0.3/-12.7 %	10.1/-17.3 %
Z+jets	-2.8/2.8 %	-2.8/2.8 %	-11.2/14.2 %	-10.0/13.1 %
Diboson	-4.5/2.2 %	-4.5/2.3 %	-13.0/13.0 %	-13.0/13.0 %

Table 17: Systematic JES down/up rate uncertainties for EMC lepton events.

7 Likelihood Function

Both combined and separate search use the same likelihood function for the template fit to the observed events. This likelihood function consists of Poisson terms for the individual bins of the fitted histogram, Gaussian constraints on the background rates, and Gaussian constraints on the strengths of systematic effects:

$$L(\beta_1, \dots, \beta_C; \delta_1, \dots, \delta_S) = \prod_{k=1}^B \frac{e^{-\mu_k} \cdot \mu_k^{n_k}}{n_k!} \cdot \prod_{j=A}^C G(\beta_j, 1.0, \Delta_j) \cdot \prod_{i=1}^S G(\delta_i, 0.0, 1.0) . \quad (8)$$

Systematic uncertainties are included as factors modifying the expectation value μ_k of events in a certain bin k , with

$$\mu_k = \sum_{j=1}^C \beta_j \cdot \hat{\nu}_j \cdot \left\{ \sum_{i=1}^S (1 + \delta_i \cdot \epsilon_{ji}) \right\} \cdot \alpha_{jk} \cdot \left\{ 1 + \sum_{i=1}^S (\delta_i \cdot \kappa_{jik}) \right\} , \quad (9)$$

$$G(\beta_j, 1.0, \Delta_j) = \frac{1}{\sqrt{2\pi} \Delta_j} \cdot \exp \left(\frac{-(\beta_j - 1.0)^2}{2 \Delta_j^2} \right) , \text{ and} \quad (10)$$

$$G(\delta_i, 0.0, 1.0) = \frac{1}{\sqrt{2\pi}} \cdot \exp \left(\frac{-\delta_i^2}{2} \right) . \quad (11)$$

The index j runs over the different event categories occurring in the likelihood function.

The predicted expectation values for the number of events of a certain event category are denoted $\hat{\nu}_j$. The free parameters in the fit are given by $\beta_j = \nu_j / \hat{\nu}_j$, i.e. the expectation values over their prediction. The normalized content of bin k of the template histogram for event category j is α_{jk} . The total number of bins is B .

In the fit, several effects causing systematic rate uncertainties are considered as given by tables 9 – 17. In this notation, the sources of systematic uncertainties carry the index i . The variation in strength of a systematic effect i is measured with the variable δ_i which constitutes an additional fit parameter and measures the strength of the systematic effect in units of one standard deviation. The relative rate uncertainties due to these sources are named ϵ_{ji} . As outlined in section 6, several sources of uncertainties influencing the template shape are taken into account. In the template distributions, the shape uncertainties are reflected by relative uncertainties in the bin content of bin k , being denominated as κ_{jik} . The values of κ_{jik} are calculated from the systematically shifted normalized template histograms α_{jik}^+ and α_{jik}^- according to

$$\kappa_{jik} = \frac{\alpha_{jik}^+ - \alpha_{jik}^-}{2 \alpha_{jk}} . \quad (12)$$

By construction the κ_{jik} satisfy the normalization condition

$$\sum_{k=1}^B \kappa_{jik} = 0 . \quad (13)$$

The systematically shifted template taking into account the shifts caused by all systematic effects with strengths $\{\delta_i\}$ is given by

$$\alpha'_{ji} = \alpha_{jk} \cdot \left\{ 1 + \sum_{i=1}^S \delta_i \cdot \kappa_{jik} \right\}. \quad (14)$$

Due to (13), the shifted histogram α'_{ji} is properly normalized:

$$\sum_{k=1}^B \alpha'_{ji} = 1. \quad (15)$$

Both normalized background rates β_j , where j runs over the background event categories, and the parameters δ_i describing the strengths of systematic excursions are constrained by Gaussian terms in the likelihood function, see (8): the background rates β_j are constrained within the relative uncertainties of the prediction, Δ_j (see table 18, while the strengths of the systematic effects δ_i are constrained to 0.0 with a standard deviation of 1.0. The single-top-quark content is measured by fitting the parameters of the likelihood function, β_j and δ_j , to the observed data. This is achieved by minimizing the negative logarithm of the likelihood function (8) with respect to these parameters using the program MINUIT [?]. In doing so, the normalized expectation values of single-top-quark events, β_1 for the combined search and β_1 and β_2 for the separate search, are allowed to assume only values greater or equal than zero to avoid unphysical results.

Using this technique, one can compute the likelihood function of the combined search as a function of β_1 only by minimizing the negative log-likelihood at a fixed value of β_1 with respect to all other variables (often called nuisance parameters). This method, often named “profiling the likelihood function”, results in a one-dimensional function, the reduced likelihood $\mathcal{L}_{\text{red}}(\beta_1)$.

process	Δ
t -ch	12.6 %
s -ch	12.4 %
single-top	12.6 %
$t\bar{t}$	12.4 %
$Wb\bar{b}+Wc\bar{c}$	30.0 %
Wc	30.0 %
Mistags	12.6 %
Z +jets	10.8 %
Diboson	1.9 %
QCD	40.0 %

Table 18: Gaussian constraints

8 Ensemble Tests

8.1 Expected Sensitivity

To compute the sensitivity of both combined and separate search, ensemble tests are used. In this context, an ensemble test consists of a set of pseudo experiments. For each pseudo experiment, first the number of events N_j of each event category is determined by drawing a random number from a Poisson distribution of a mean $\hat{\nu}_j$. As a result, the pseudo experiment features a total number of $\sum N_j$ events.

In a second step, N_j random numbers are drawn from the template distributions of the neural network output for all considered event categories displayed figure 18– 19 for the combined search and figures 20– 23, and 24– 25 for the separate search, respectively. Those random numbers are filled in a histogram which constitutes the neural network output distribution of a particular pseudo experiment. For the combined search, two ensemble tests are performed: one with single-top-quark events included at the predicted standard-model rate and one without any single-top-quark events.

8.2 Expected Significance

To compute the significance of a potentially observed signal, a hypothesis test is performed, considering two hypotheses. The first one, the null hypothesis H_0 , assumes that the single-top-quark cross section is zero ($\beta_1 = 0$ for the combined search; $\beta_1 = \beta_2 = 0$ for the separate search). The second one, H_1 , assumes that the single-top-quark production cross-section is the one predicted by the standard model ($\beta_1 = 1$ for the combined search; $\beta_1 = \beta_2 = 1$ for the separate search). The objective of both analyses is to observe single-top-quark events, that means to reject the null hypothesis H_0 . The hypothesis test for the combined neural network search is based on the Q -value,

$$Q = -2 (\ln L_{\text{red}}(\beta_1 = 1) - \ln L_{\text{red}}(\beta_1 = 0)) , \quad (16)$$

where $L_{\text{red}}(\beta_1 = 1)$ is the value of the reduced likelihood function at the standard-model prediction and $L_{\text{red}}(\beta_1 = 0)$ is the value of the reduced likelihood function for a single-top-quark cross section of zero. Using the two ensemble tests, the respective Q -value distributions are determined. The resulting distributions are shown in figure 41 for the combined search.

In order to quantify the probability for H_0 to be correct, the p -value is defined. Assuming that the value Q_0 is observed in a particular experiment, the p -value is given by

$$p(Q_0) = \frac{1}{I_q} \cdot \int_{-\infty}^{Q_0} q_0(Q') dQ' , \quad (17)$$

where q_0 is the distribution of Q -values for the null hypothesis H_0 and

$$I_q = \int_{-\infty}^{+\infty} q_0(Q') dQ' . \quad (18)$$

To quantify the sensitivity of the analysis, the expected p -value $\hat{p} = p(\tilde{Q}_1)$ is defined, where \tilde{Q}_1 is the median of the Q -value distribution q_1 for the hypothesis H_1 . The meaning of \hat{p} is the following: Under the assumption that H_1 is correct one expects to observe

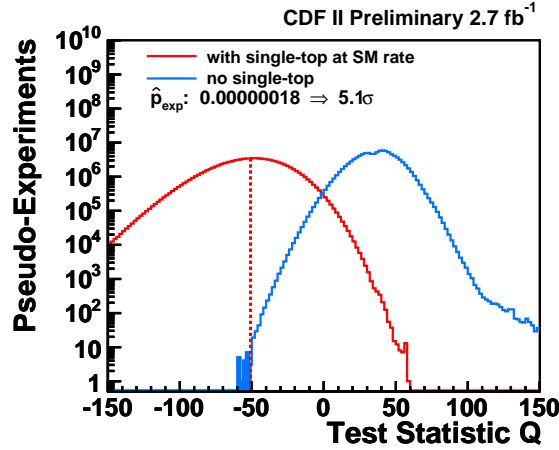


Figure 41: Distributions of Q -values for two ensemble tests, one with single-top-quark events present at the expected standard-model rate, one without any single-top-quark events for the combined search.

$p < \hat{p}$ with a probability of 50%. For the combined search, $\hat{p} = 0.00000018$ is found, which corresponds to an expected significance of 5.1σ . The value includes all rate and shape systematics. The found \hat{p} -value can be interpreted as follows: assuming the predicted single-top-quark production cross-section, the expectation is, with a probability of 50%, to see at least that many single-top-quark events that the observed excess over the background corresponds to a background fluctuation of 5.1σ in case of the combined search.

9 Application to Observed Events

After the expected sensitivity has been determined, the neural networks are applied to observed events. At first, the output distributions of observed events are compared to the expected distributions. Finally, the templates are fitted to the observed distributions to determine the single-top-quark cross sections.

9.1 Comparison to Expectation

Due to the shapes of the template distributions, the bins with the highest output values are the most interesting ones: here, the ratio of signal to background events should be the largest. The predicted and measured output distribution of all 8 channels used in the combined search are depicted in figure ???. In figure 44 the distributions of all four neural networks are added together.

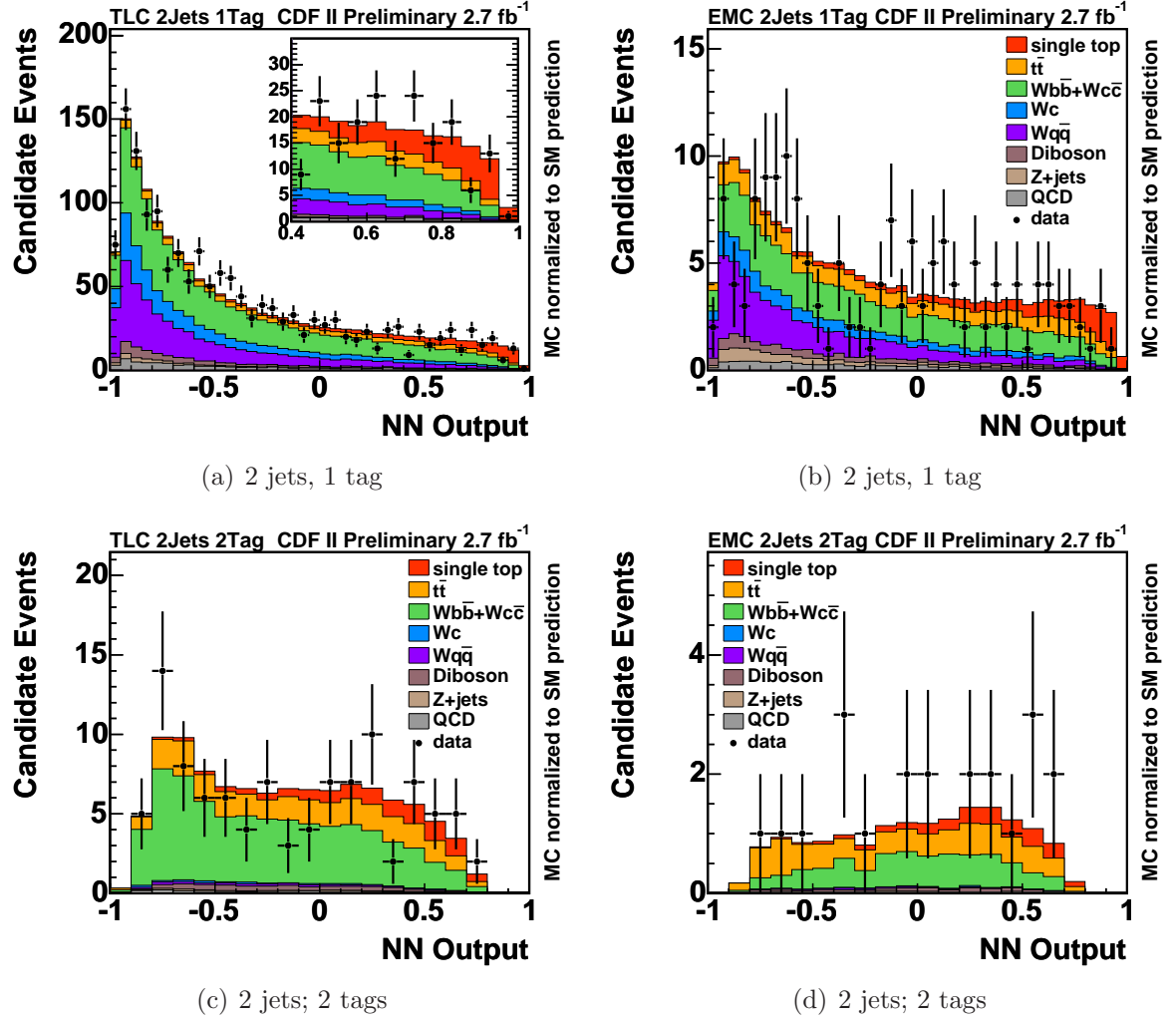


Figure 42: The predicted and measured distributions of the combined search for the t -channel neural network in the 2-jet bin with 1 b -tag (top) and the s -channel neural network in the 2-jet bin with 2 b -tags (bottom) for TLC (left) and EMC (right) leptons. The output of t - and s -channel events is added with a ratio corresponding to the standard-model prediction.

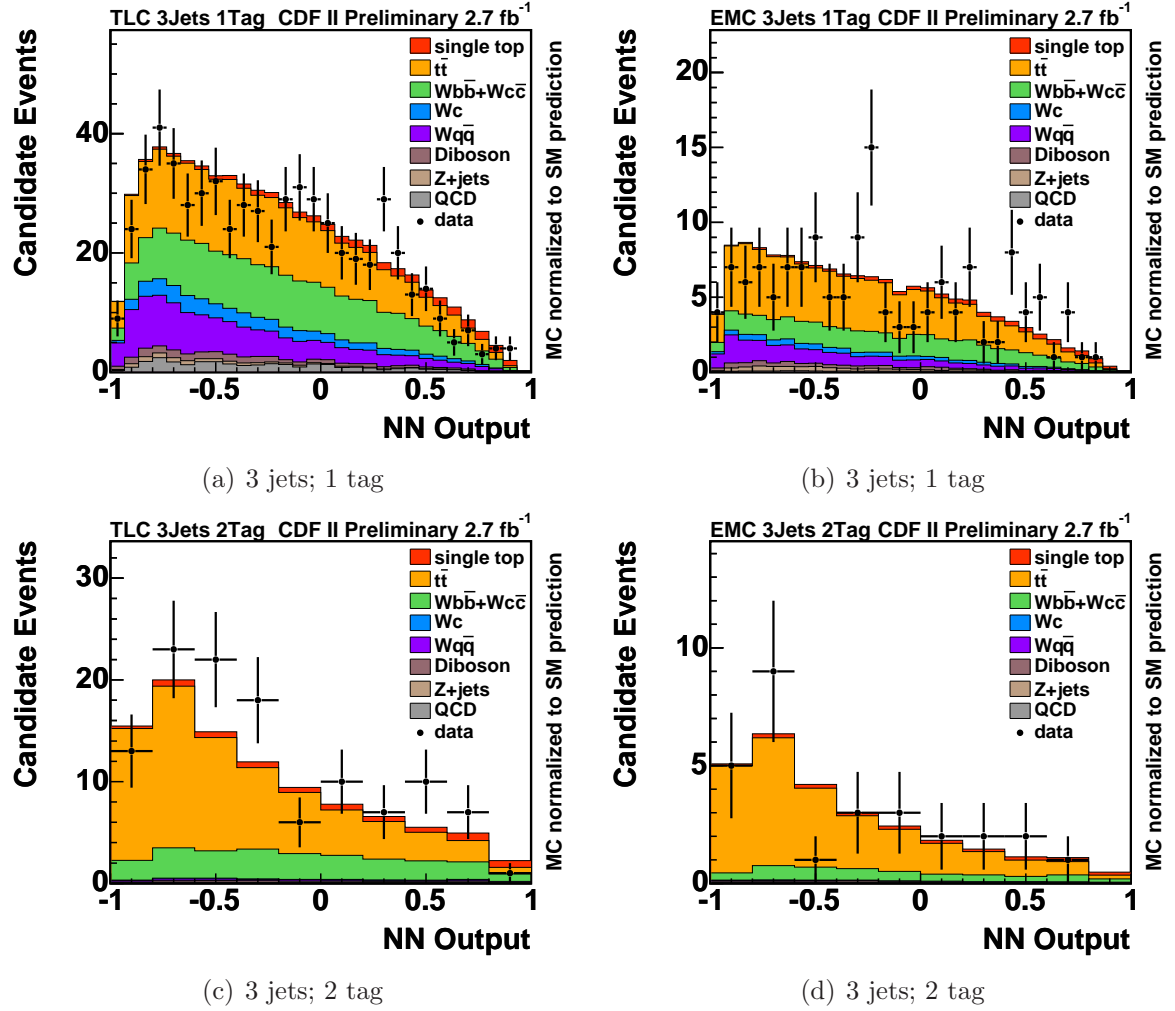


Figure 43: The predicted and measured distributions of the combined search for the t -channel neural network in the 3-jet bin with 1 b -tag (top) and in the 3-jet bin with 2 b -tags (bottom) for TLC (left) and EMC (right) leptons. The output of t - and s -channel events is added with a ratio corresponding to the standard-model prediction.

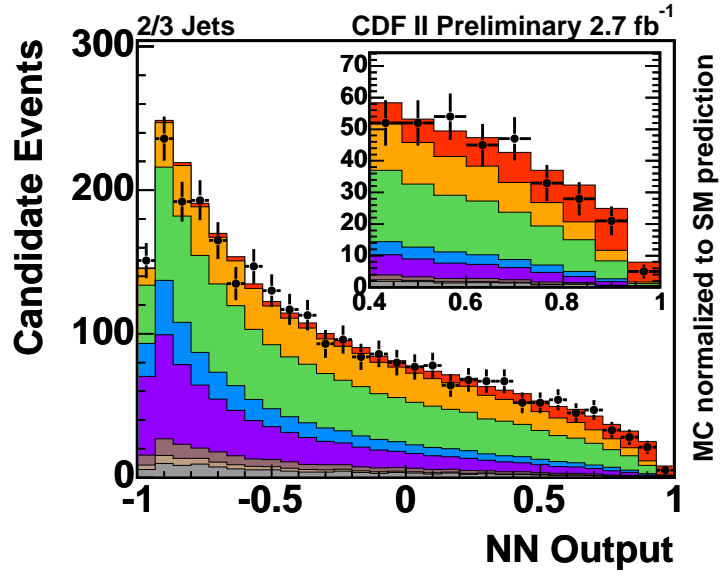


Figure 44: The predicted and measured distributions of all 8 channels used in the combined search. The output of t - and s -channel events is added with a ratio corresponding to the standard-model prediction.

9.2 Fit Results

9.2.1 Combined Search

The likelihood fit to the neural network output for the combined search yields a cross section for single top quark production of $2.1^{+0.7}_{-0.6}$ pb. The observed Q -value is -27.2, yielding an observed p -value of 0.00006790 corresponding to an observed significance of 3.8σ . Figure 45, comparing the observed Q -value with the expectation.

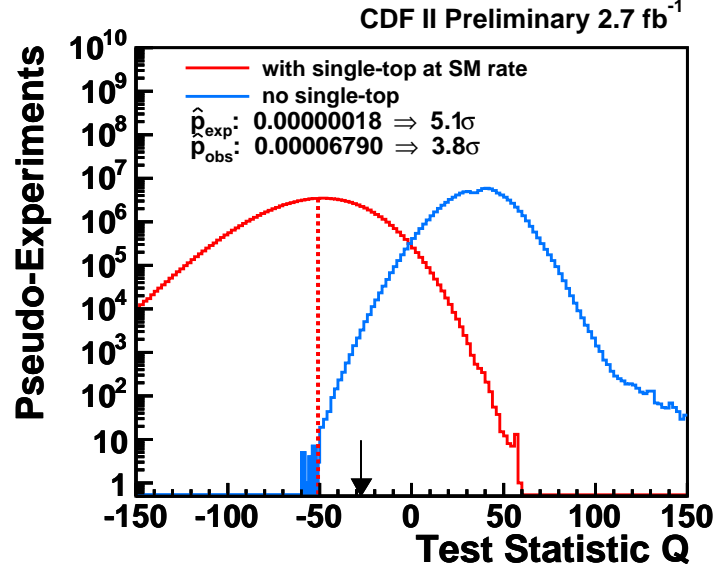


Figure 45: Comparison of observed Q -value to the expectation in the combined search.

9.2.2 Separate Search

For the separate search, the fit yields the cross sections $\sigma_t = 0.7_{-0.5}^{+0.6}(\text{stat.} + \text{syst.})$ pb for t -channel and $\sigma_s = 2.1_{-0.7}^{+0.7}(\text{stat.} + \text{syst.})$ pb for s -channel, respectively. The fit result is depicted in figure 46 showing the difference between the logarithm of the reduced likelihood function and its minimum in the plane of single-top-quark s -channel versus t -channel cross sections. Negative cross section values are physically meaningless and therefore not allowed. The minimum represents the best fit values and is indicated by the black dot. The error bars quote the 1σ , 2σ , and 3σ uncertainties ($\Delta\ln(L)$ of 0.50, 2.0, and 4.5) on the fitted t - and s -channel cross sections. The true values of both cross sections have a probability of 68.3%, 95.5%, and 99.7% to be found in the region comprised by the corresponding contours ($\Delta\ln(L)$ of 1.15, 3.09, and 5.92). The value predicted by the standard model, within its uncertainties, is illustrated by the blue rectangle.

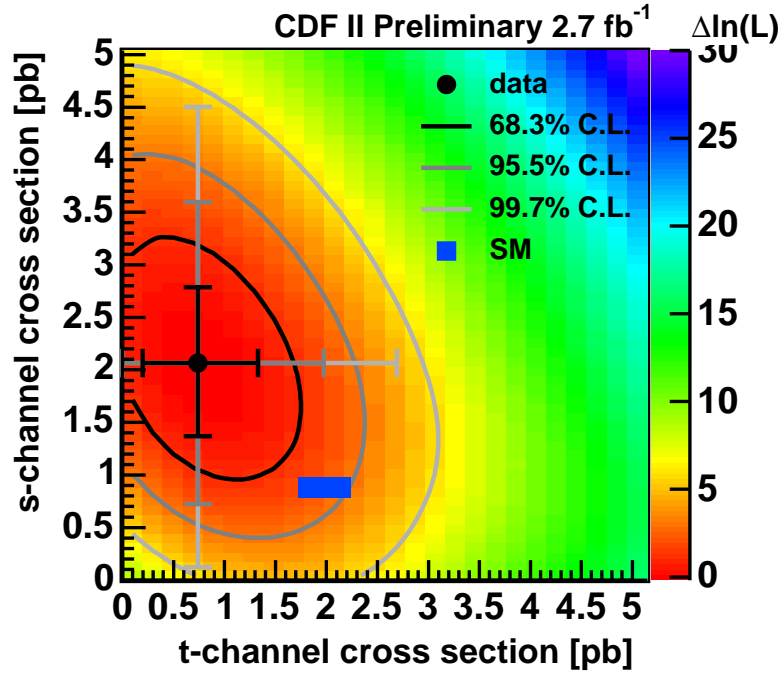


Figure 46: Logarithmic likelihood function of the separate neural network search. The 1, 2, and 3 standard deviation error bars for the one-dimensional and contours for the two-dimensional uncertainties are overlaid.

A Input Variables

A.1 Variables for the 2jet 2tag bin neural network

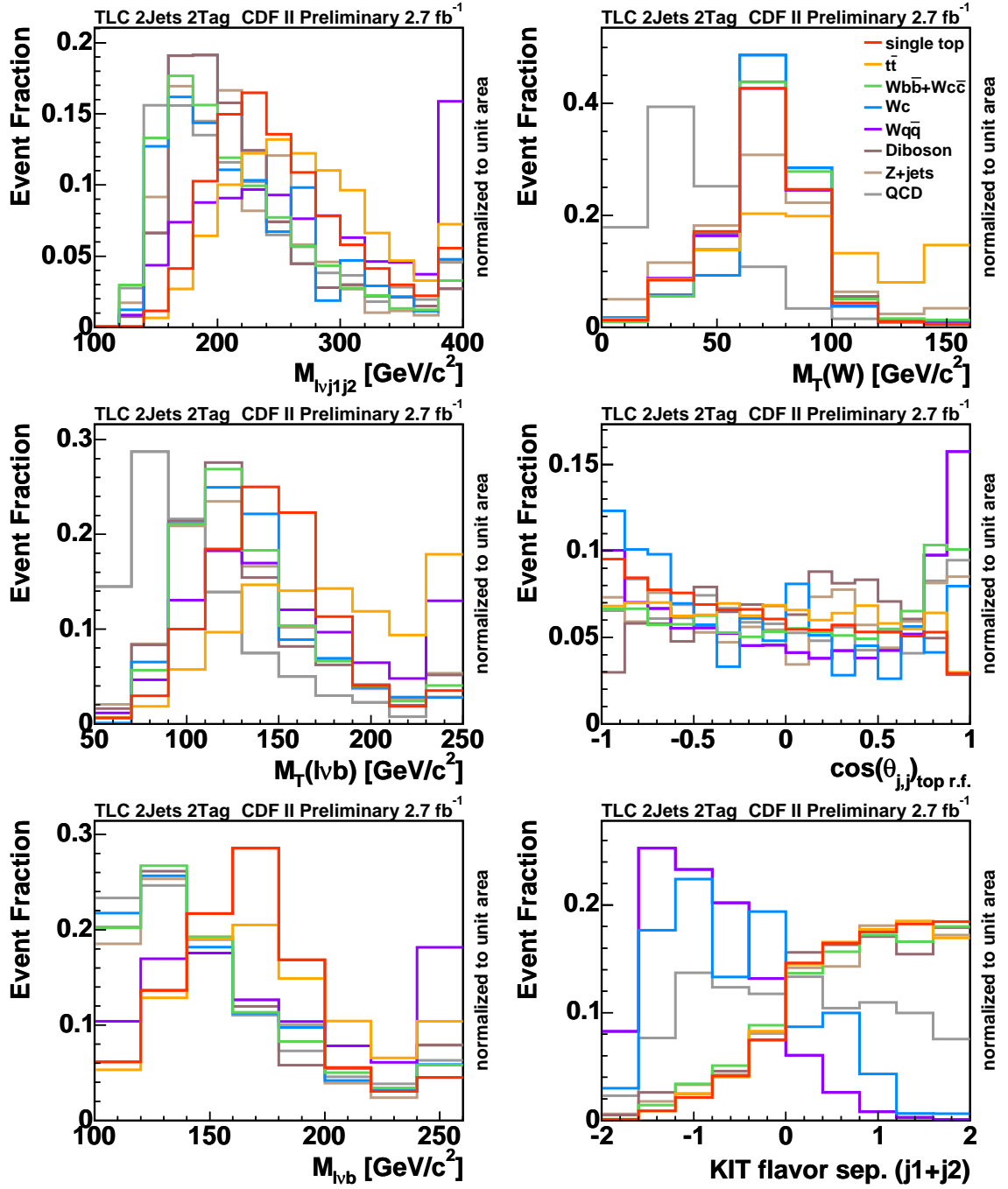


Figure 47: Shape comparison for input variables used in the neural network trained for single top s-channel requiring 2 jets and 2 tags for TLC leptons.

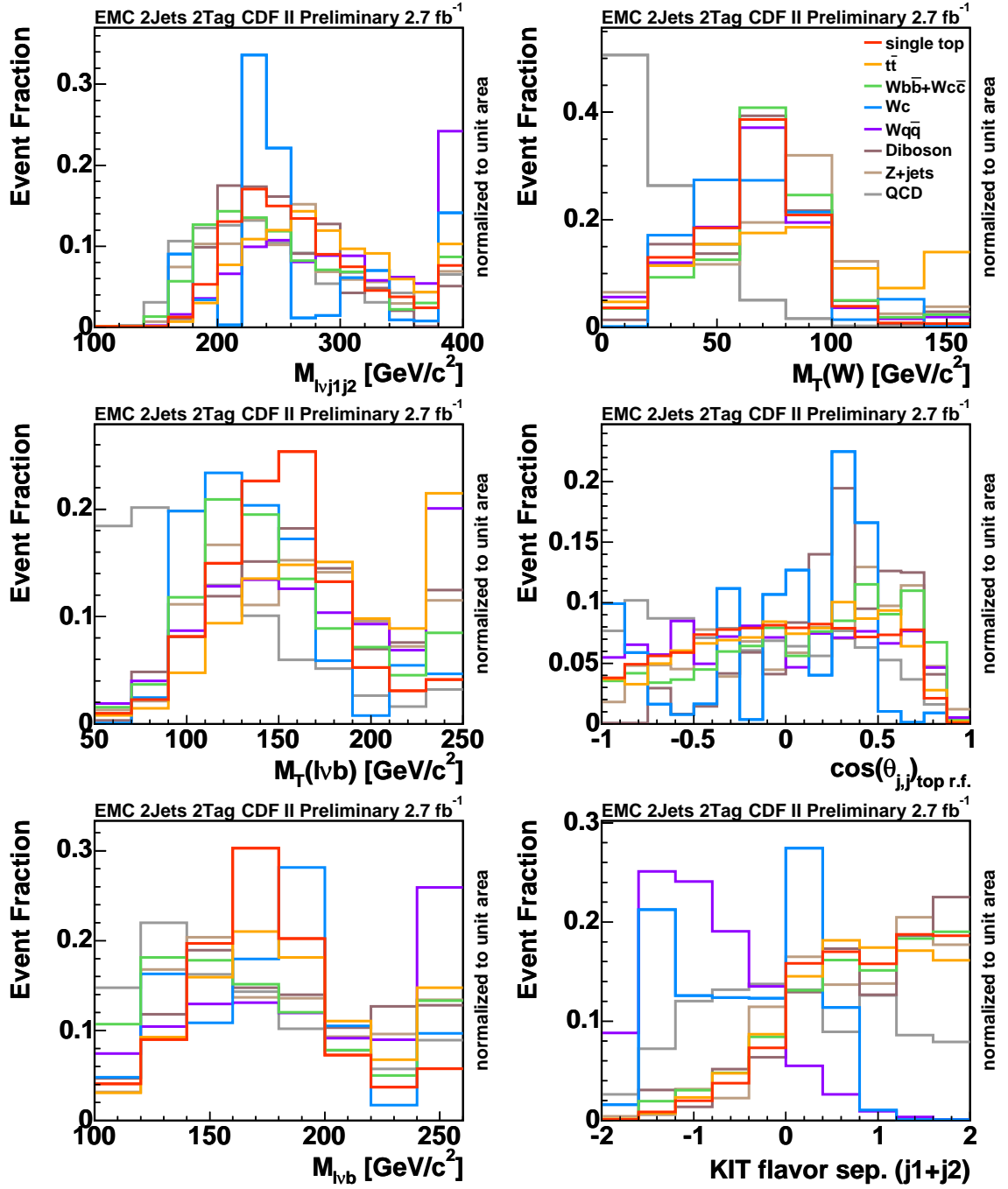


Figure 48: Shape comparison for input variables used in the neural network trained for single top s-channel requiring 2 jets and 2 tags for EMC leptons.

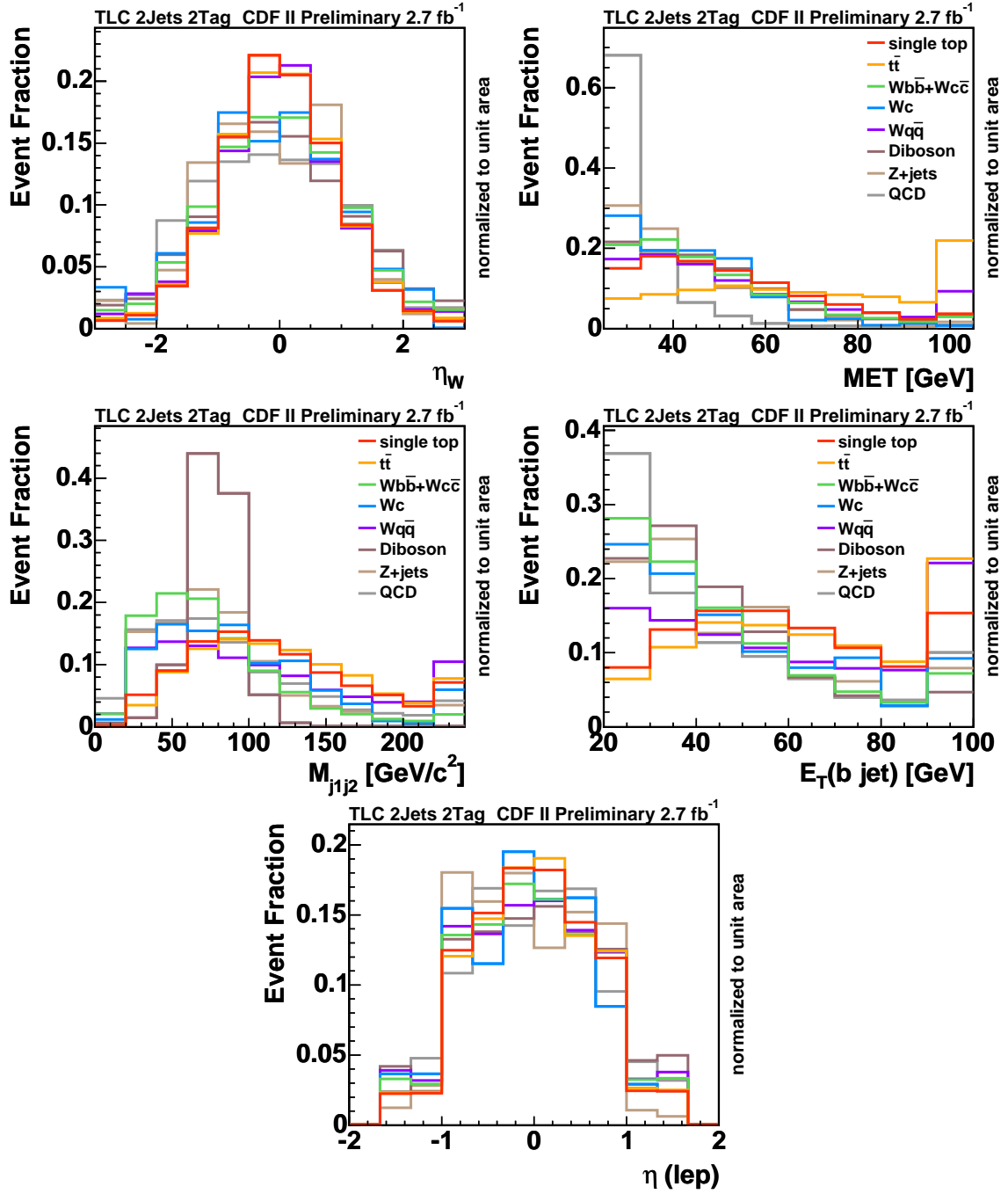


Figure 49: Shape comparison for input variables used in the neural network trained for single top s-channel requiring 2 jets and 2 tags for TLC leptons.

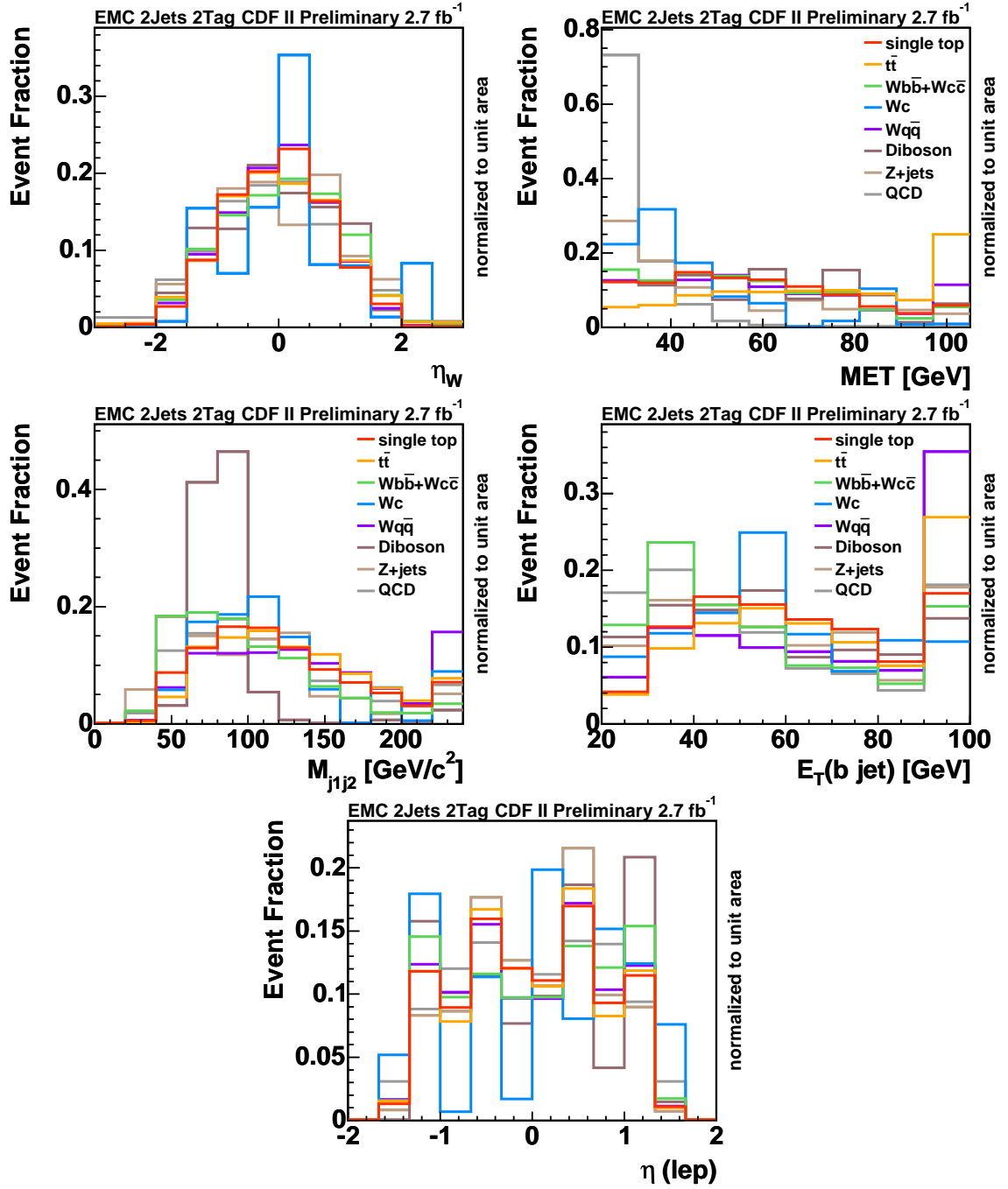


Figure 50: Shape comparison for input variables used in the neural network trained for single top s-channel requiring 2 jets and 2 tags for EMC leptons.

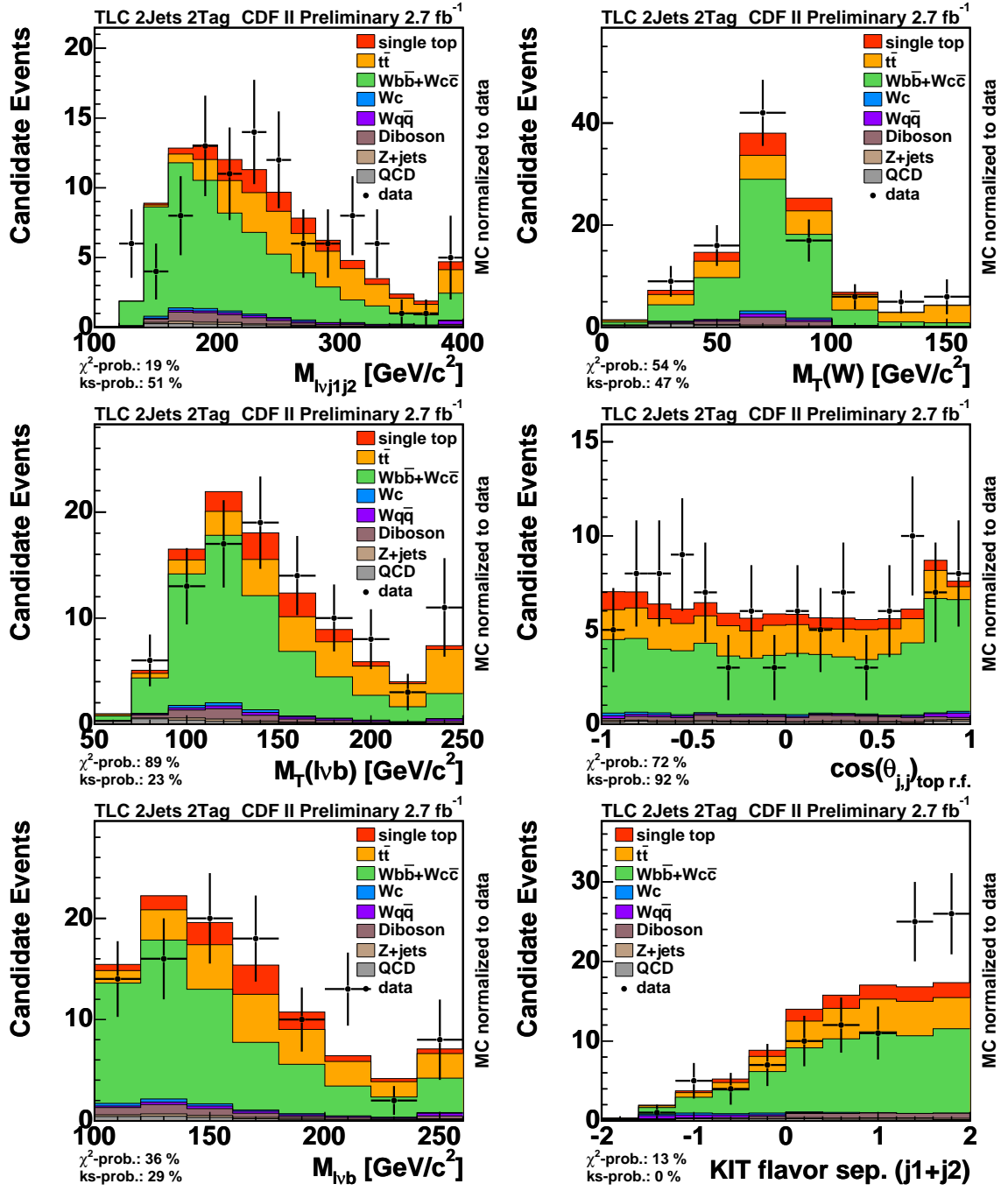


Figure 51: Data MC comparison for input variables used in the neural network trained for single top s-channel requiring 2 jets and 2 tags for TLC leptons.

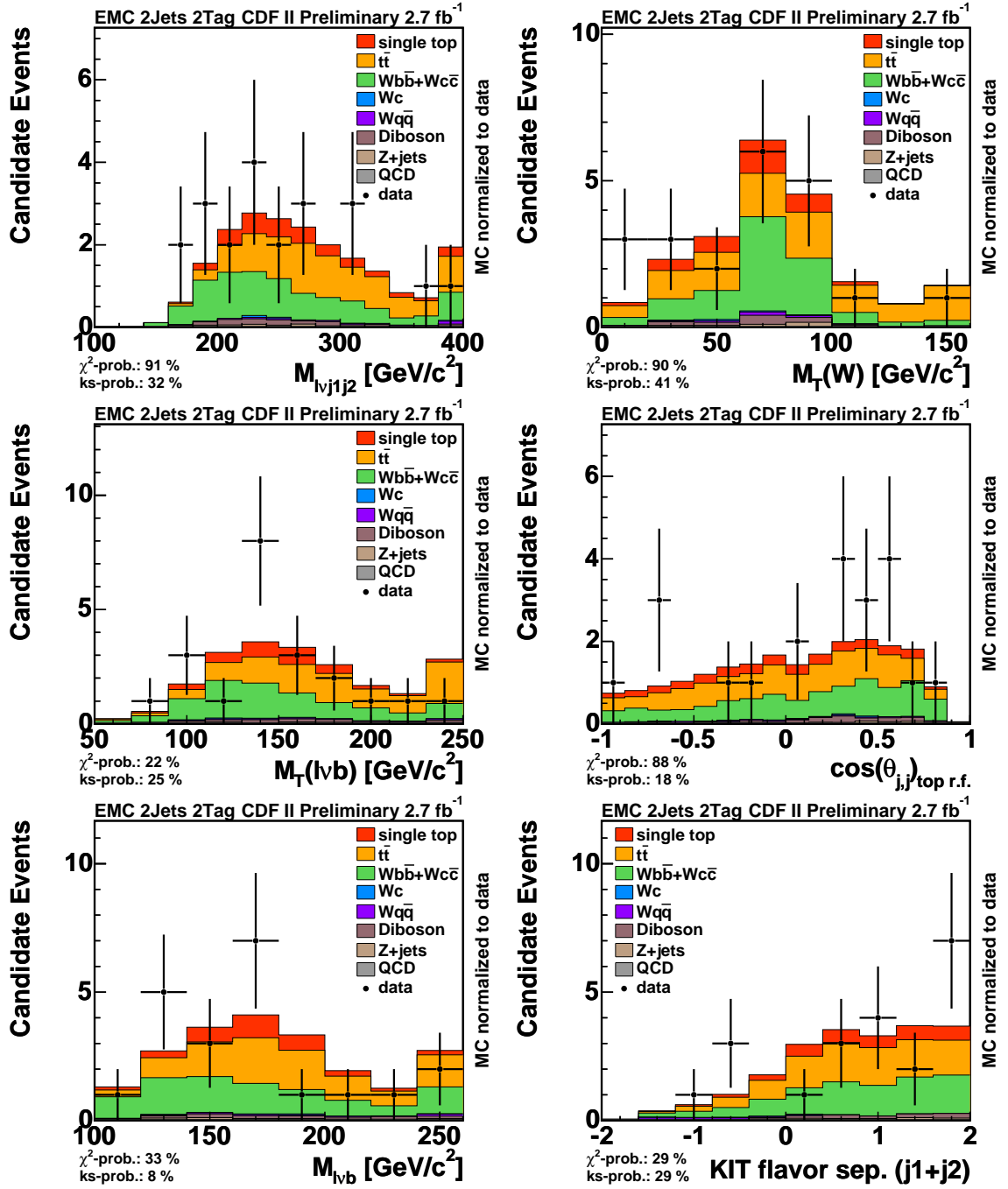


Figure 52: Data MC comparison for input variables used in the neural network trained for single top s-channel requiring 2 jets and 2 tags for EMC leptons.

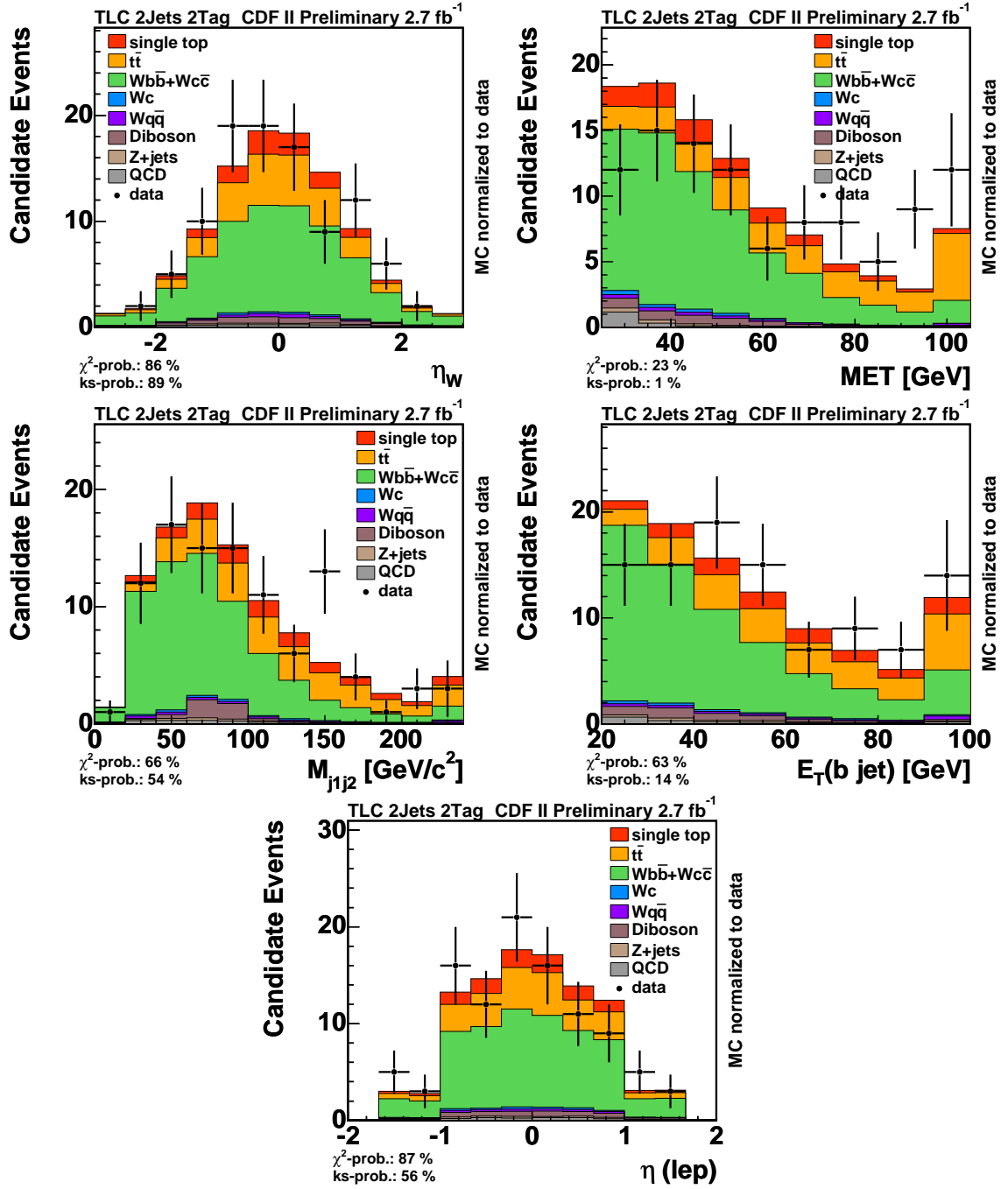


Figure 53: Data MC comparison for input variables used in the neural network trained for single top s-channel requiring 2 jets and 2 tags for TLC leptons.

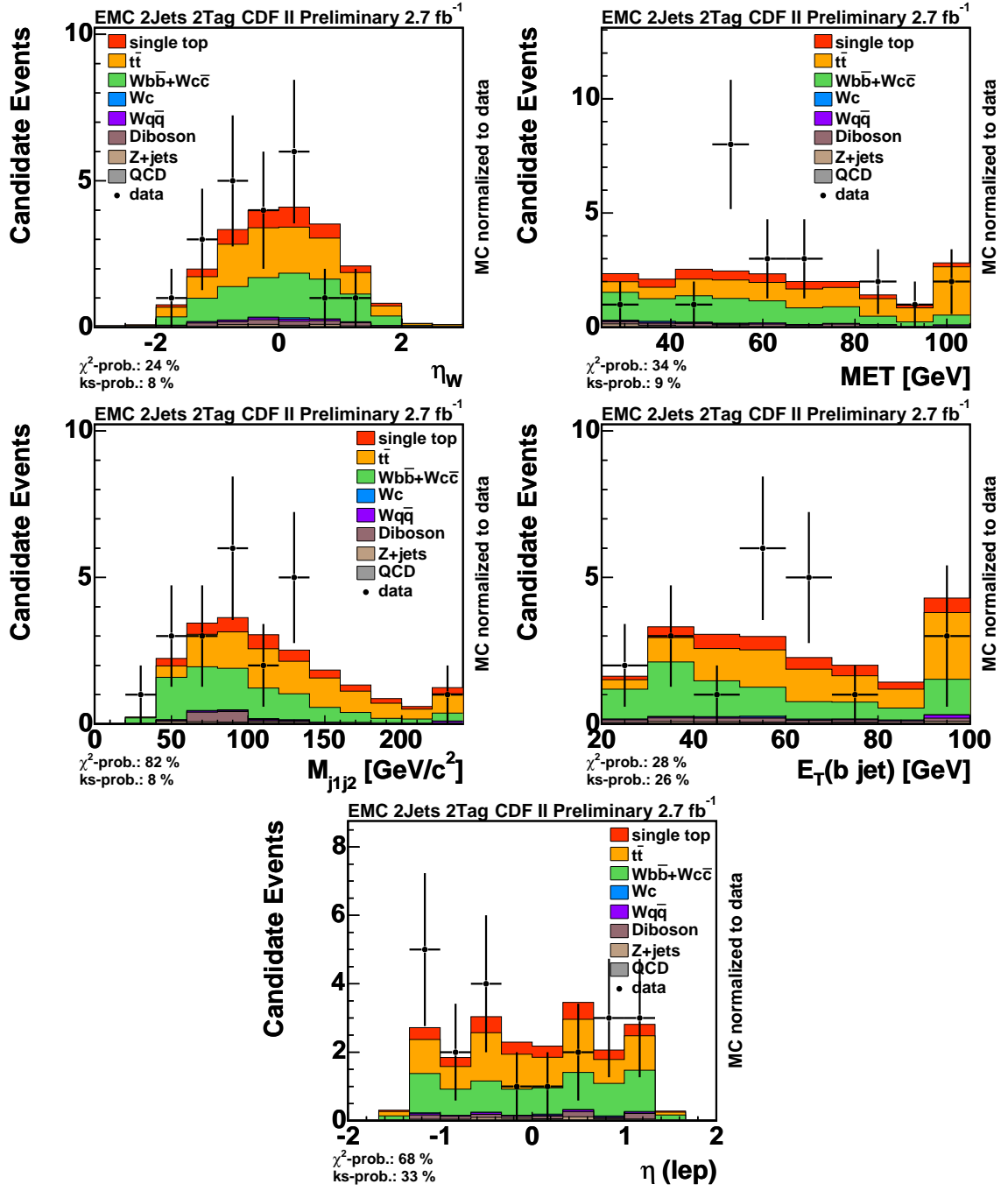


Figure 54: Data MC comparison for input variables used in the neural network trained for single top s-channel requiring 2 jets and 2 tags for EMC leptons.

A.2 Variables for the 3jet 1tag bin neural network

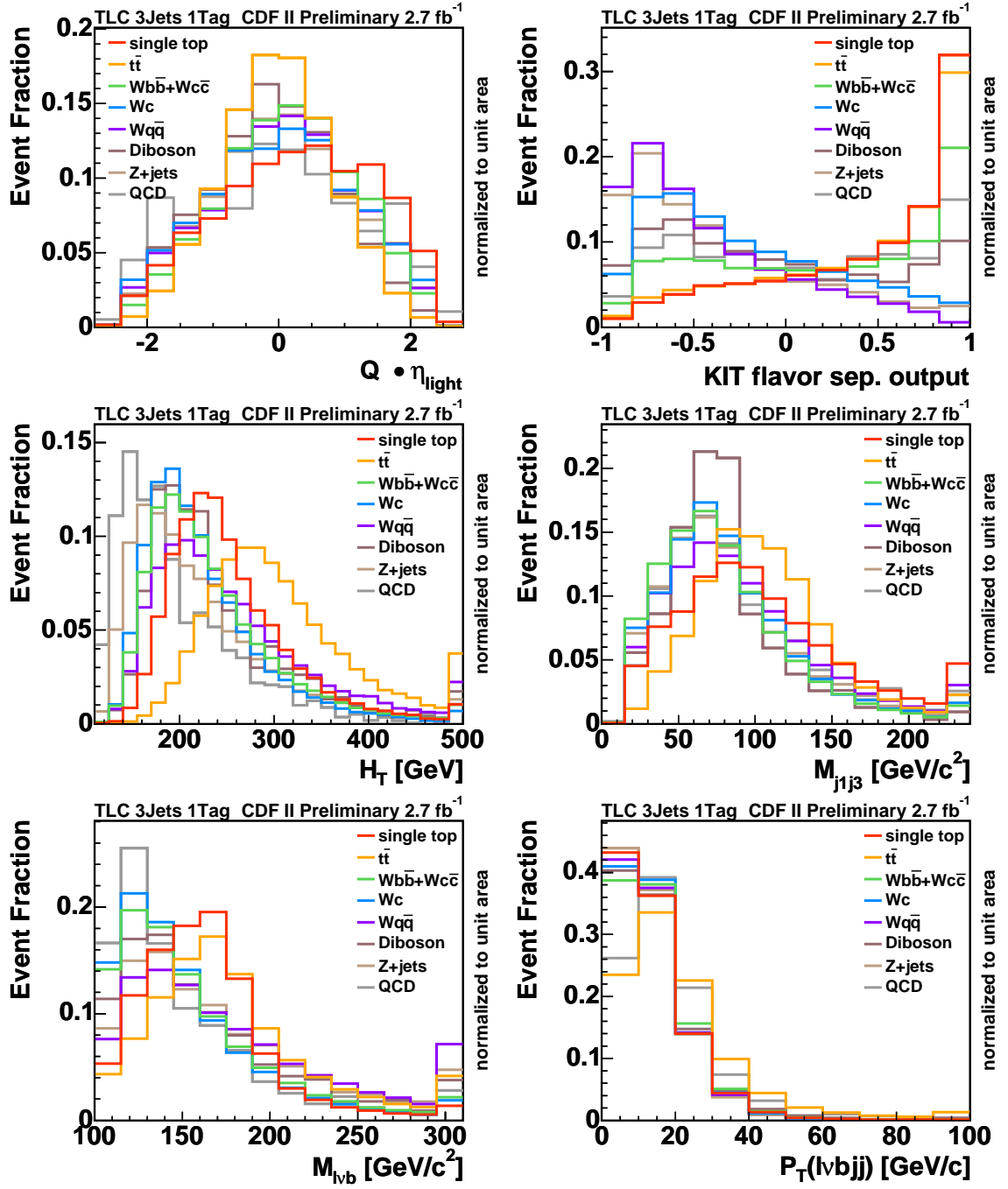


Figure 55: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for TLC leptons.

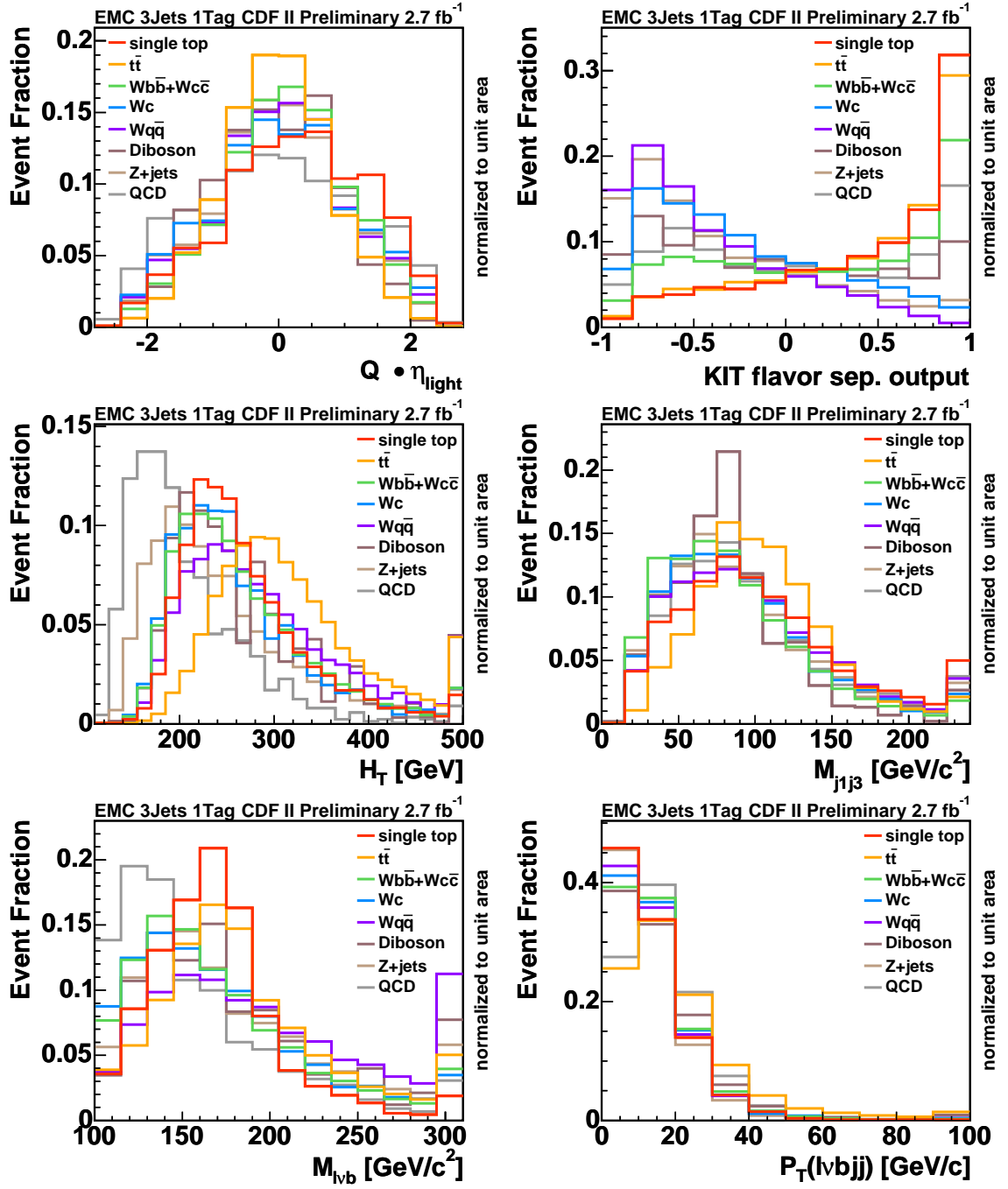


Figure 56: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for EMC leptons.

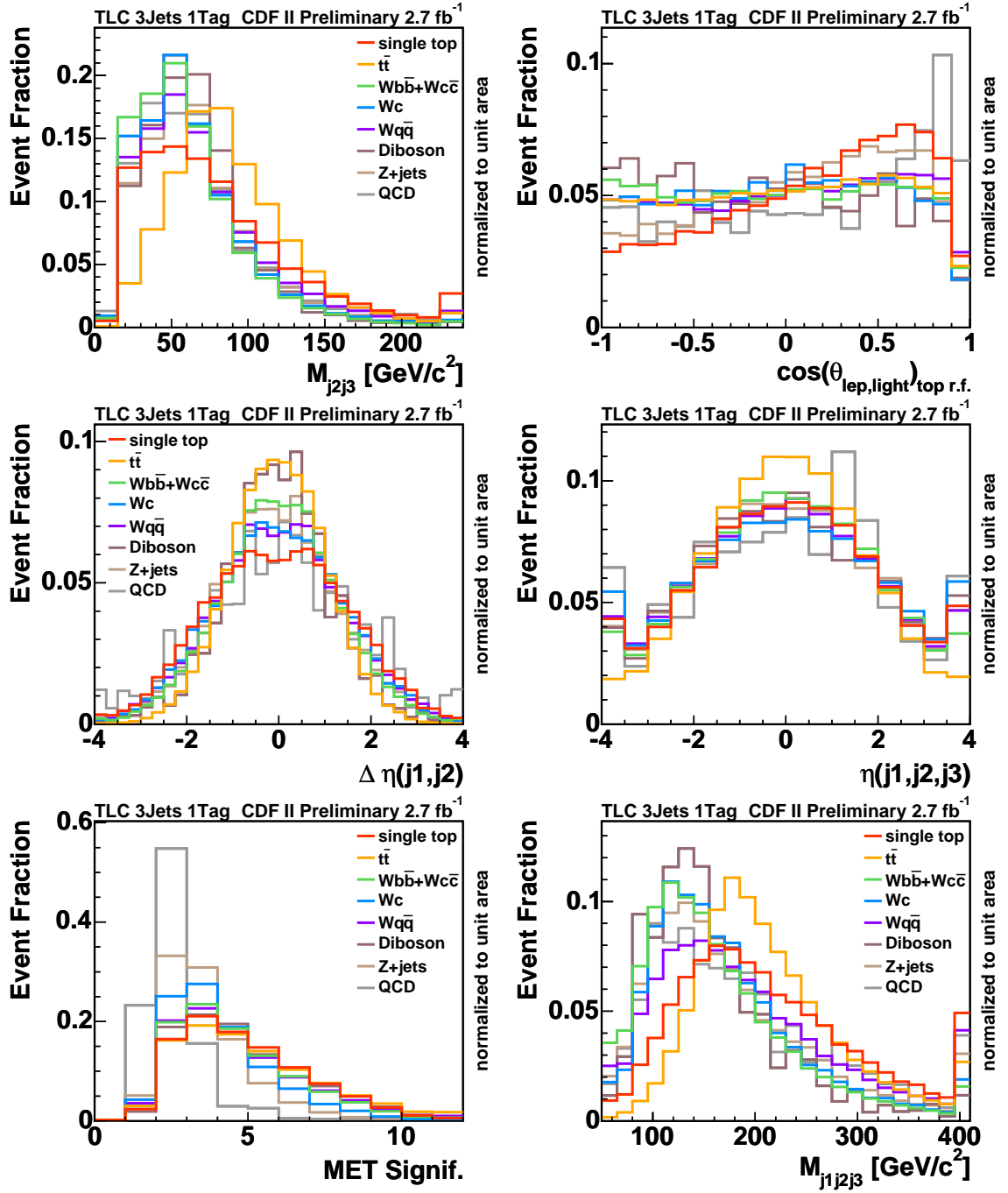


Figure 57: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for TLC leptons.

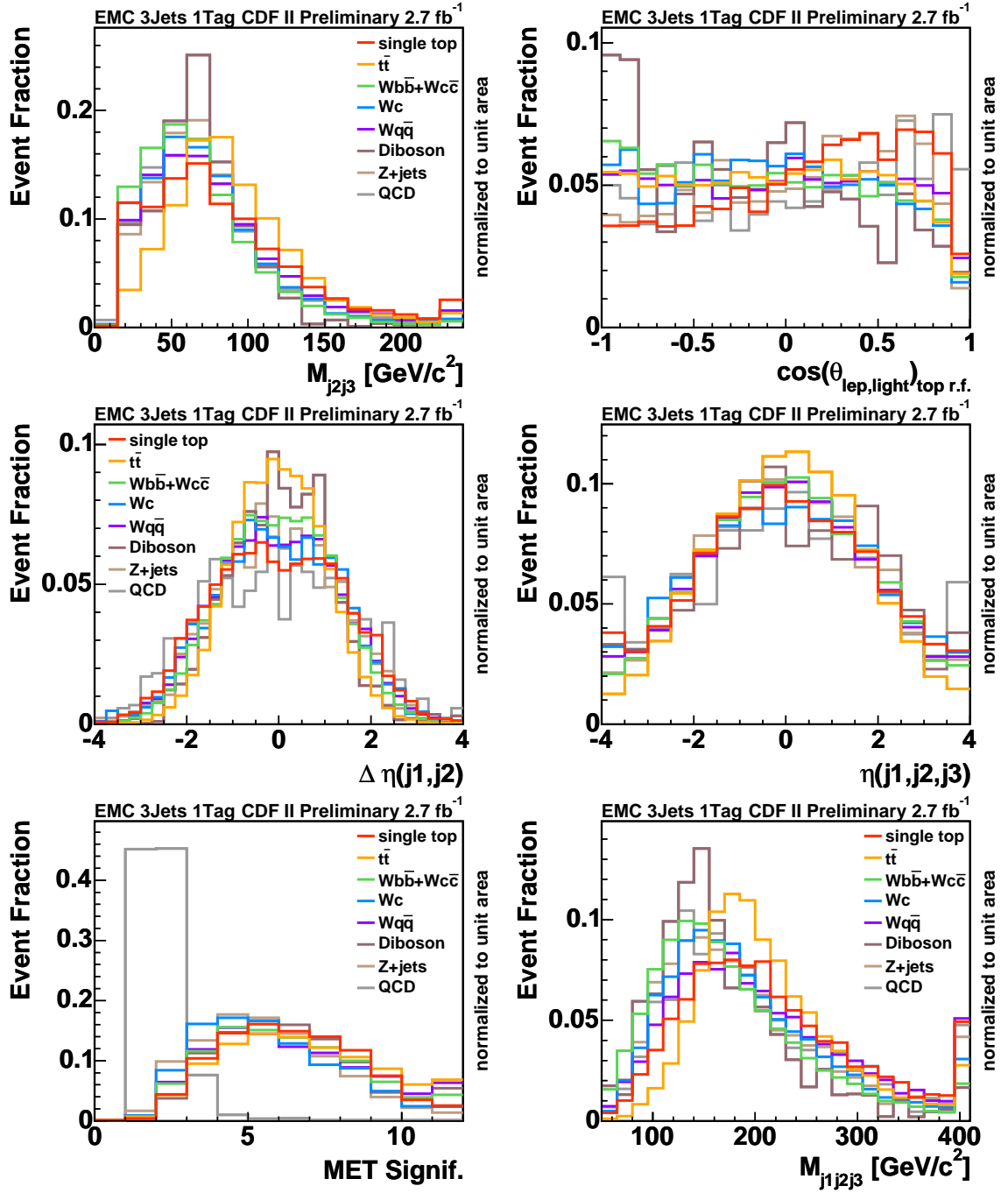


Figure 58: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for EMC leptons.

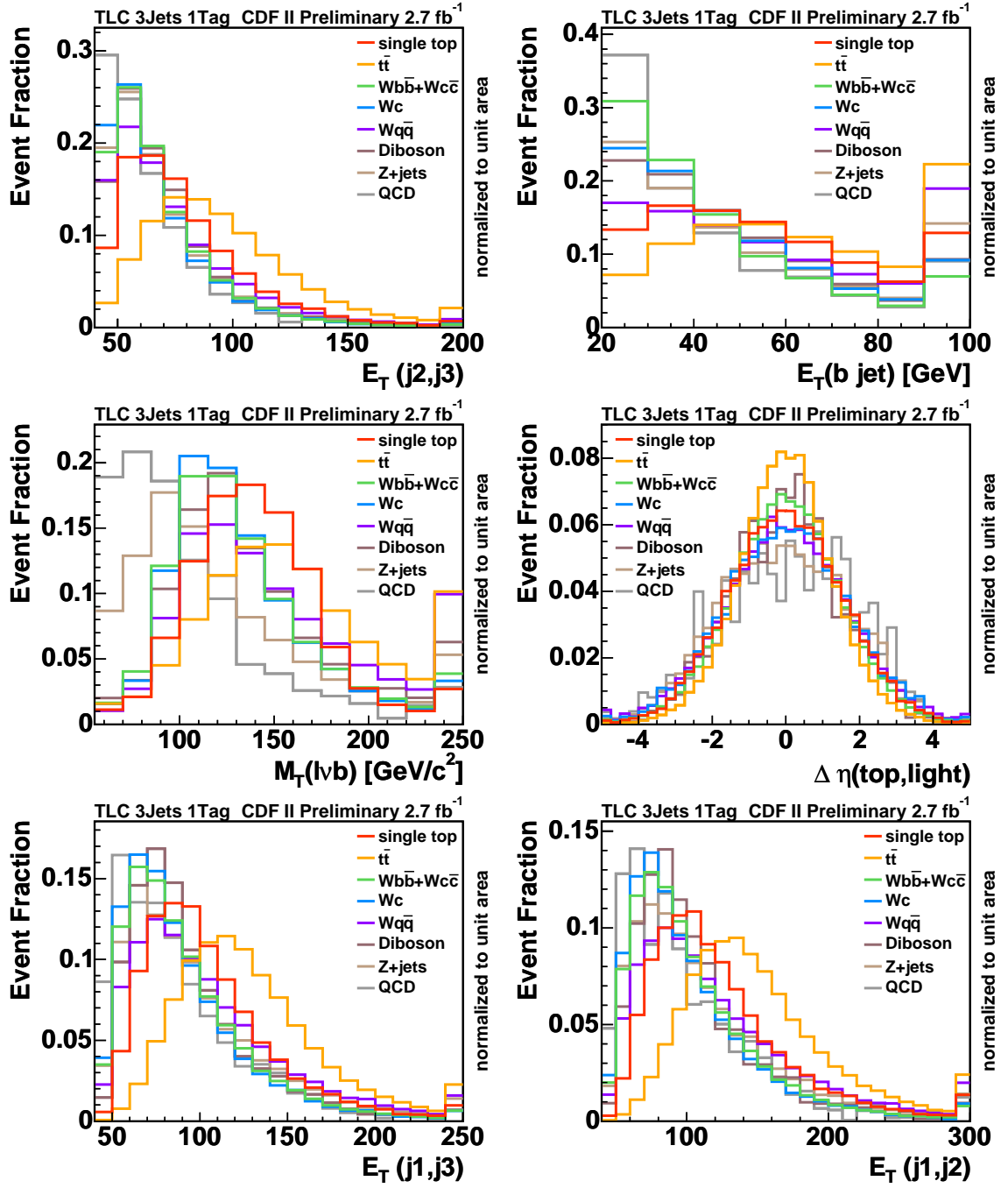


Figure 59: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for TLC leptons.

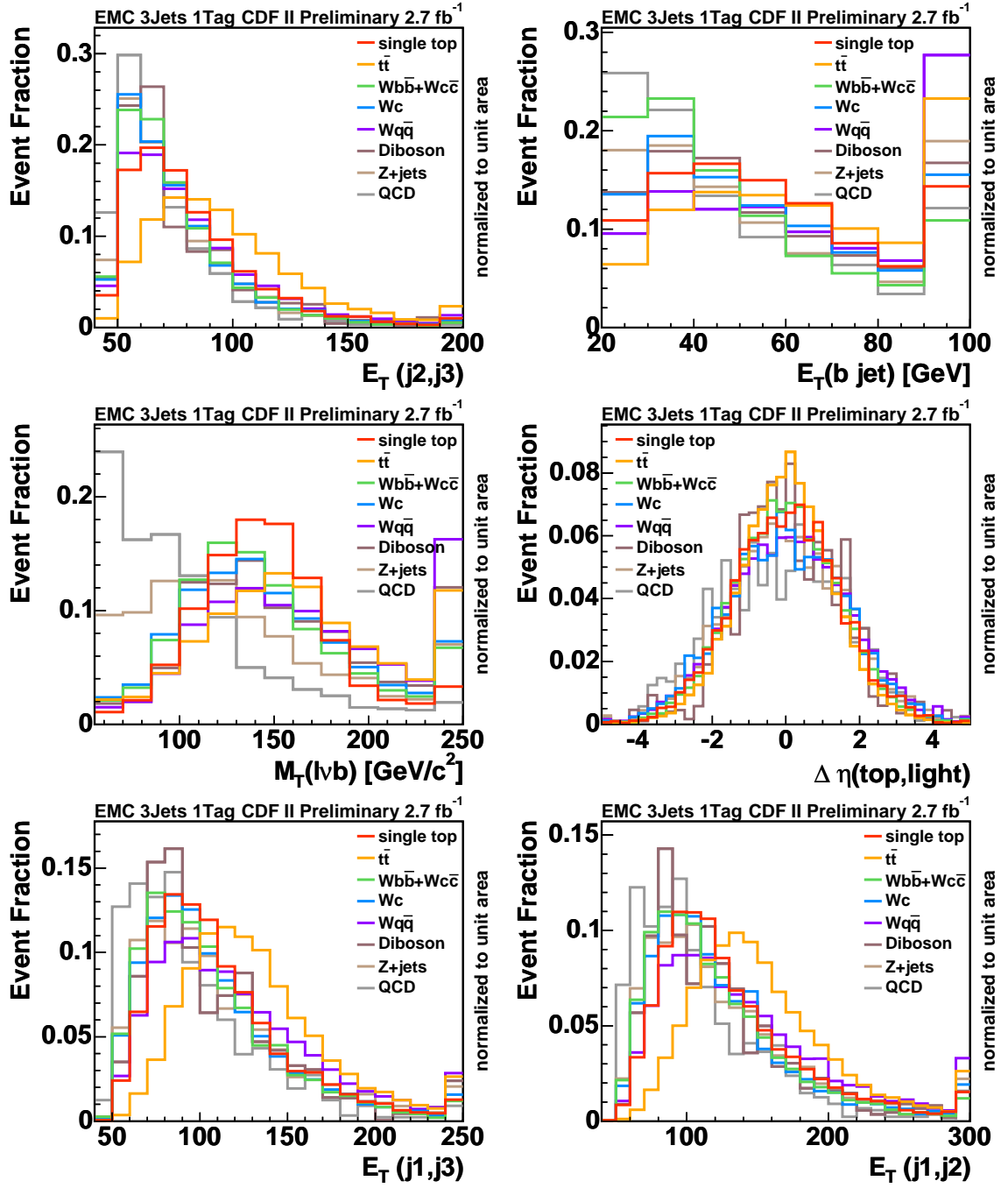


Figure 60: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for EMC leptons.

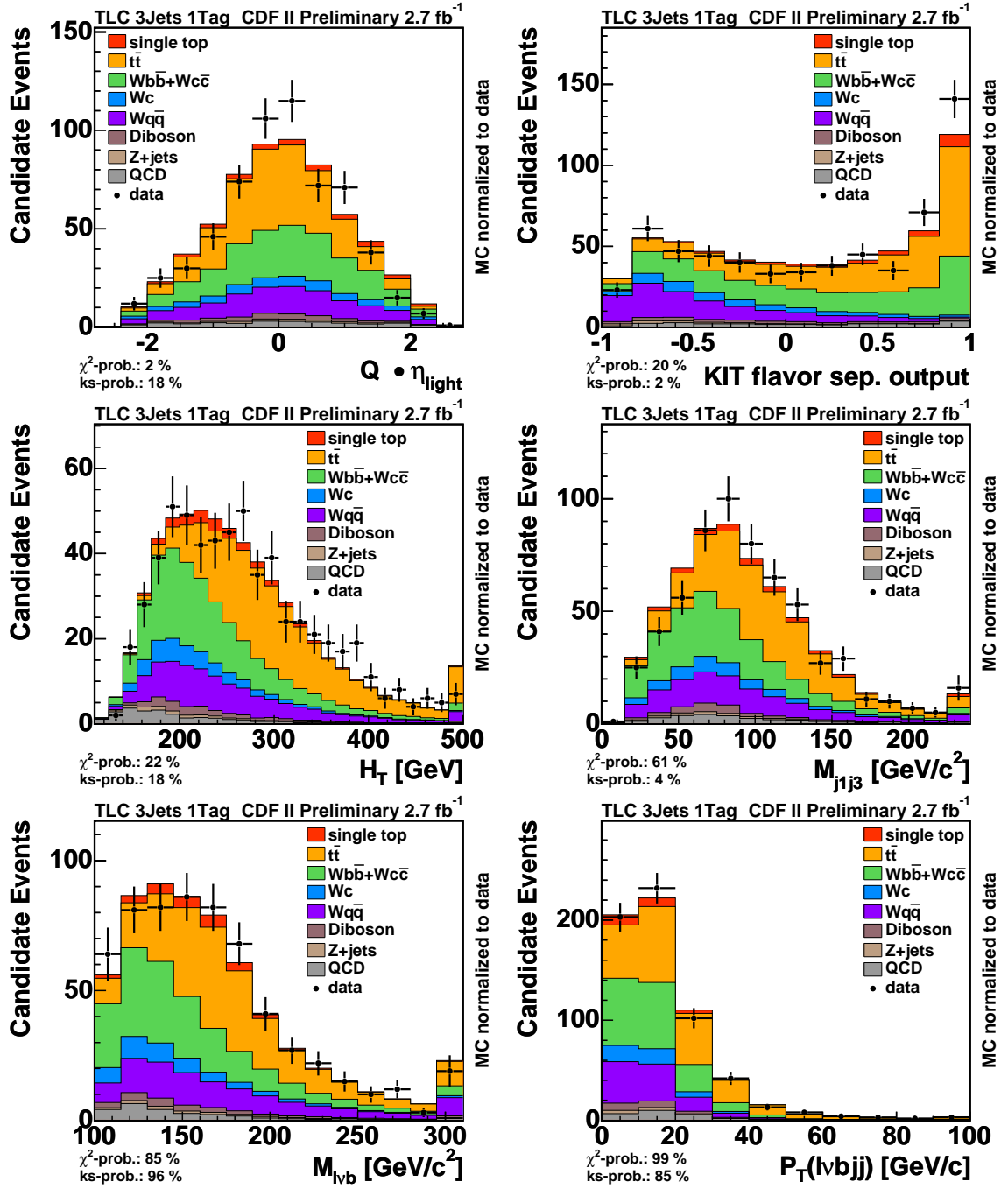


Figure 61: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for TLC leptons.

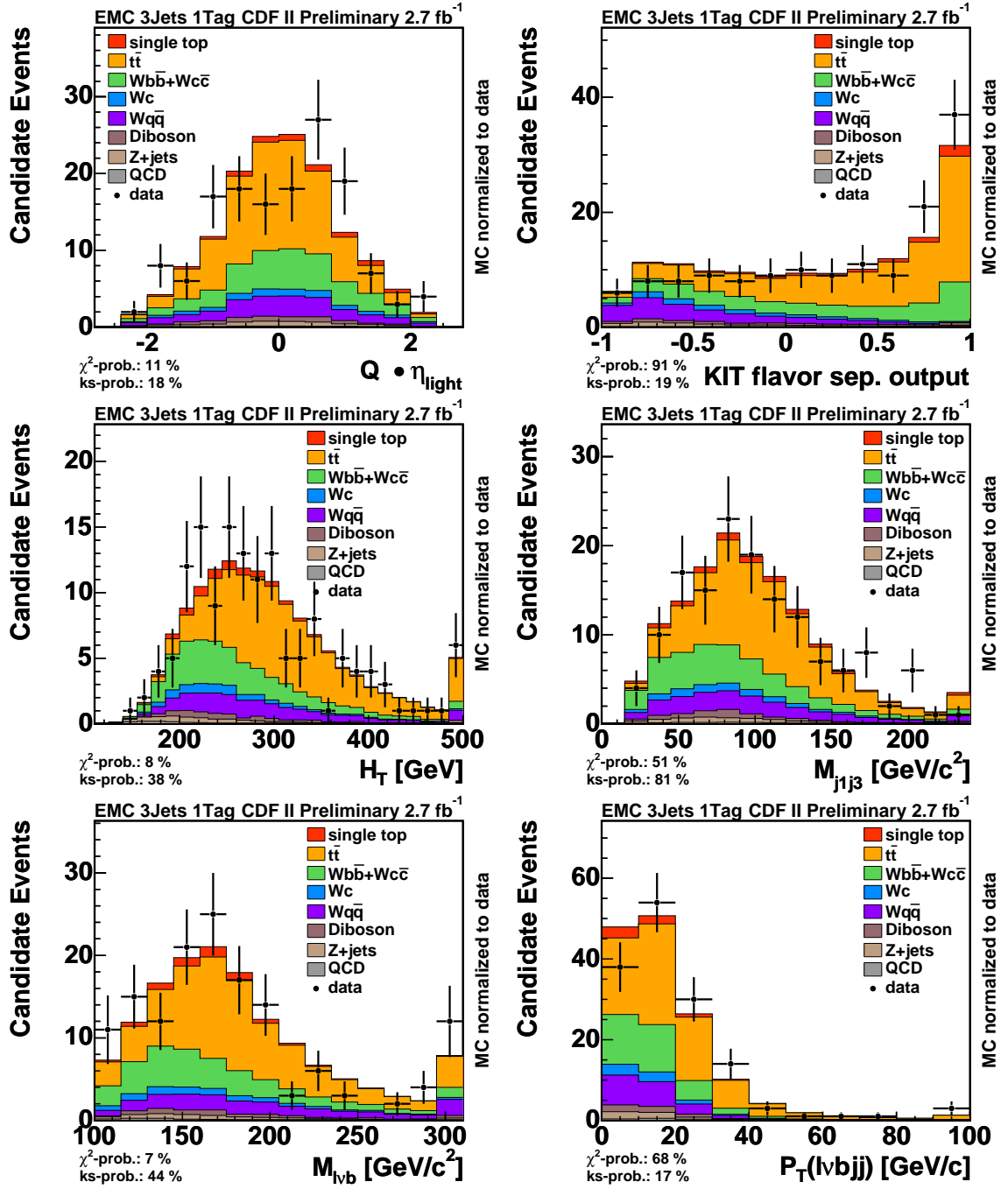


Figure 62: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for EMC leptons.

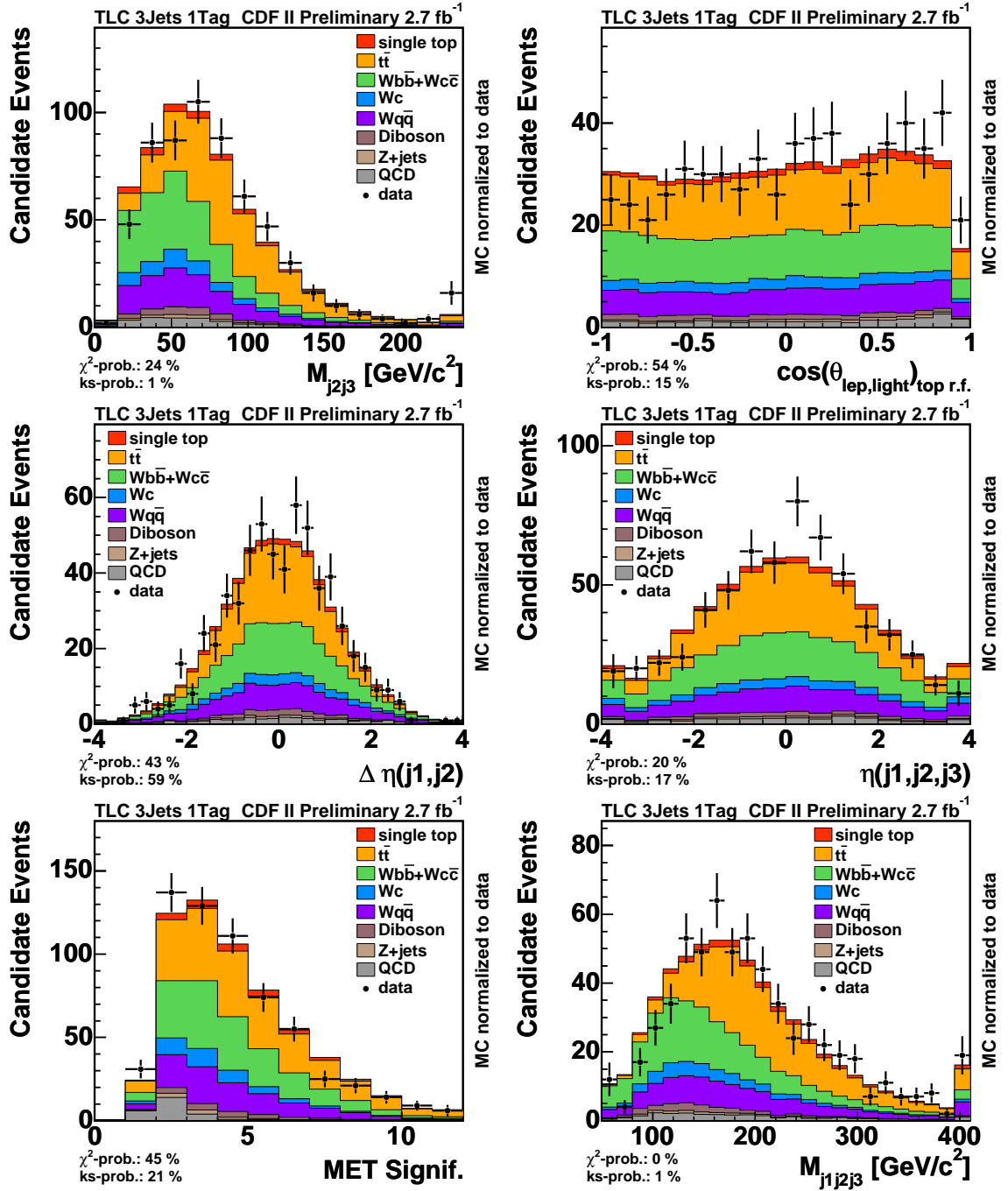


Figure 63: Shape comparison for input variables used in the t-channel neural network requiering 3 jets and 1 tag for TLC leptons.

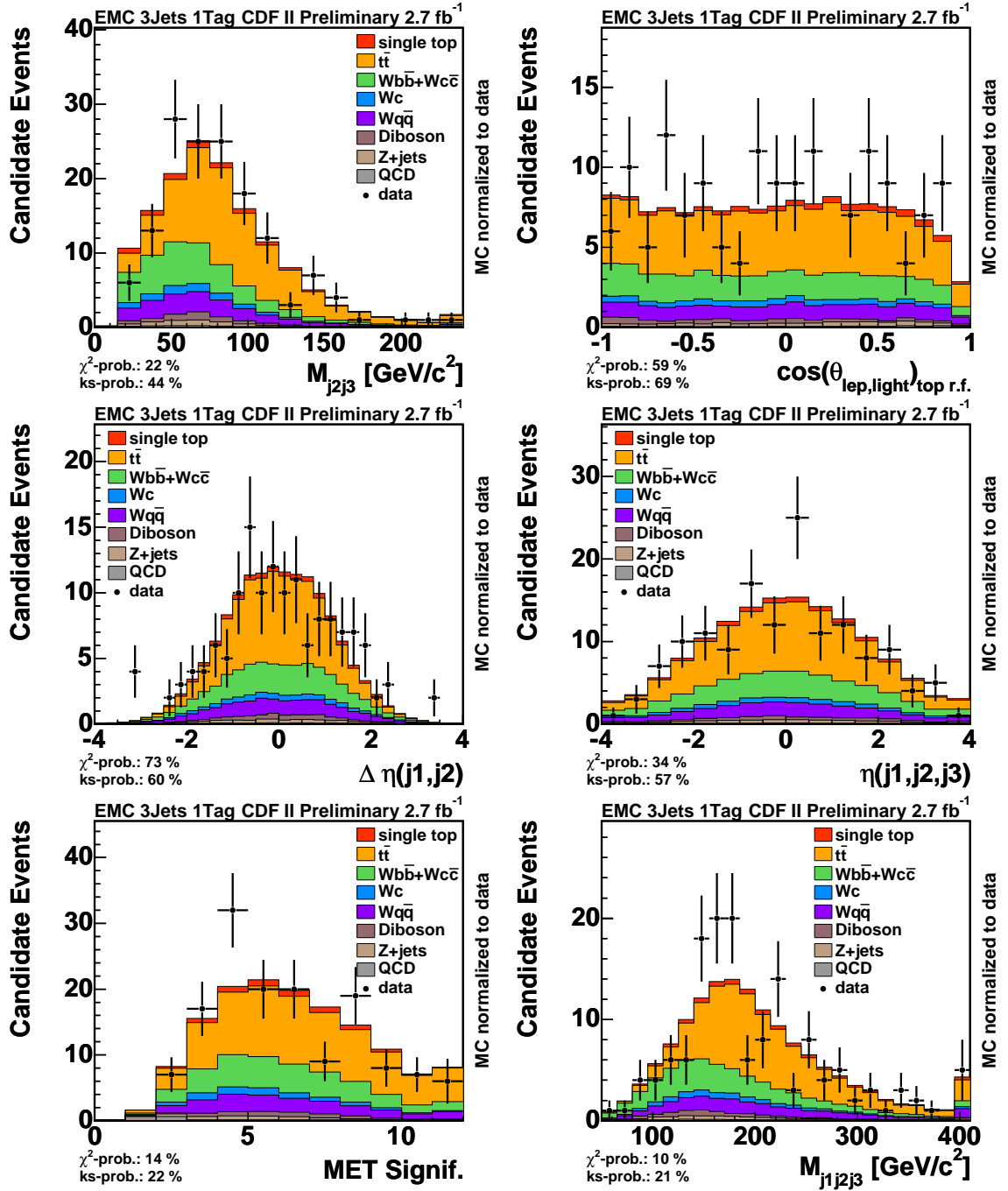


Figure 64: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for EMC leptons.

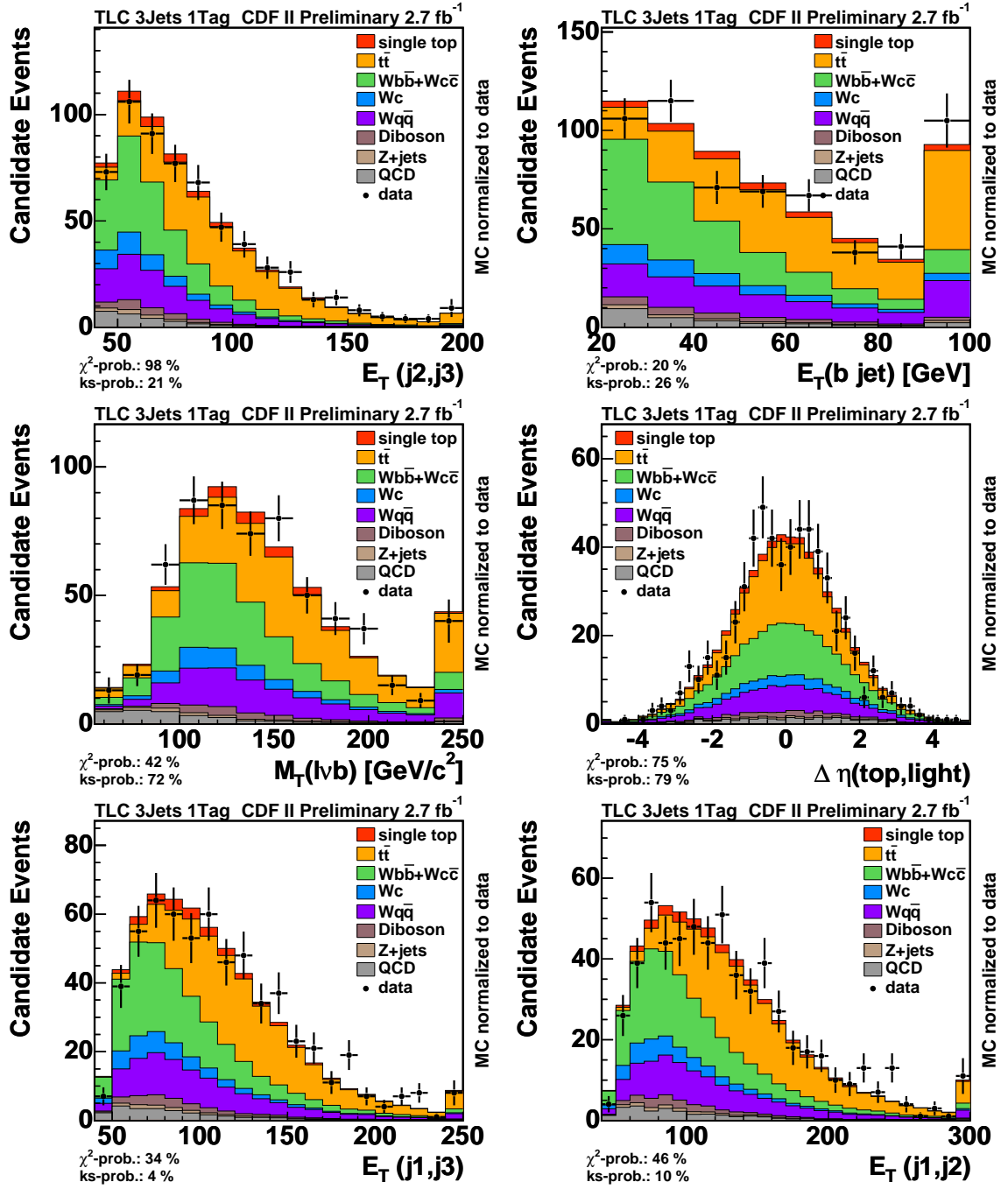


Figure 65: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for TLC leptons.

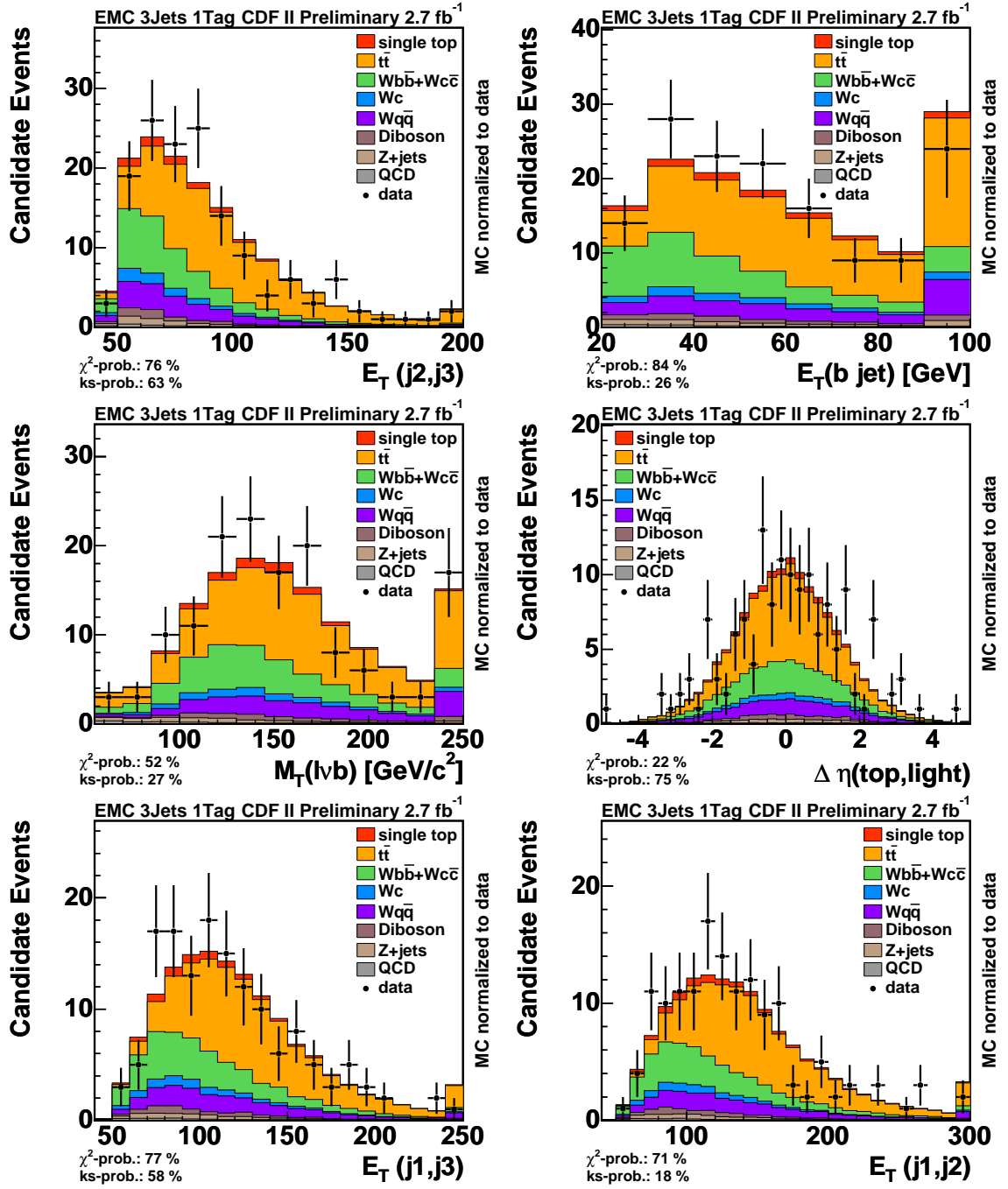


Figure 66: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 1 tag for EMC leptons.

A.3 Variables for the 3jet 2tag bin neural network

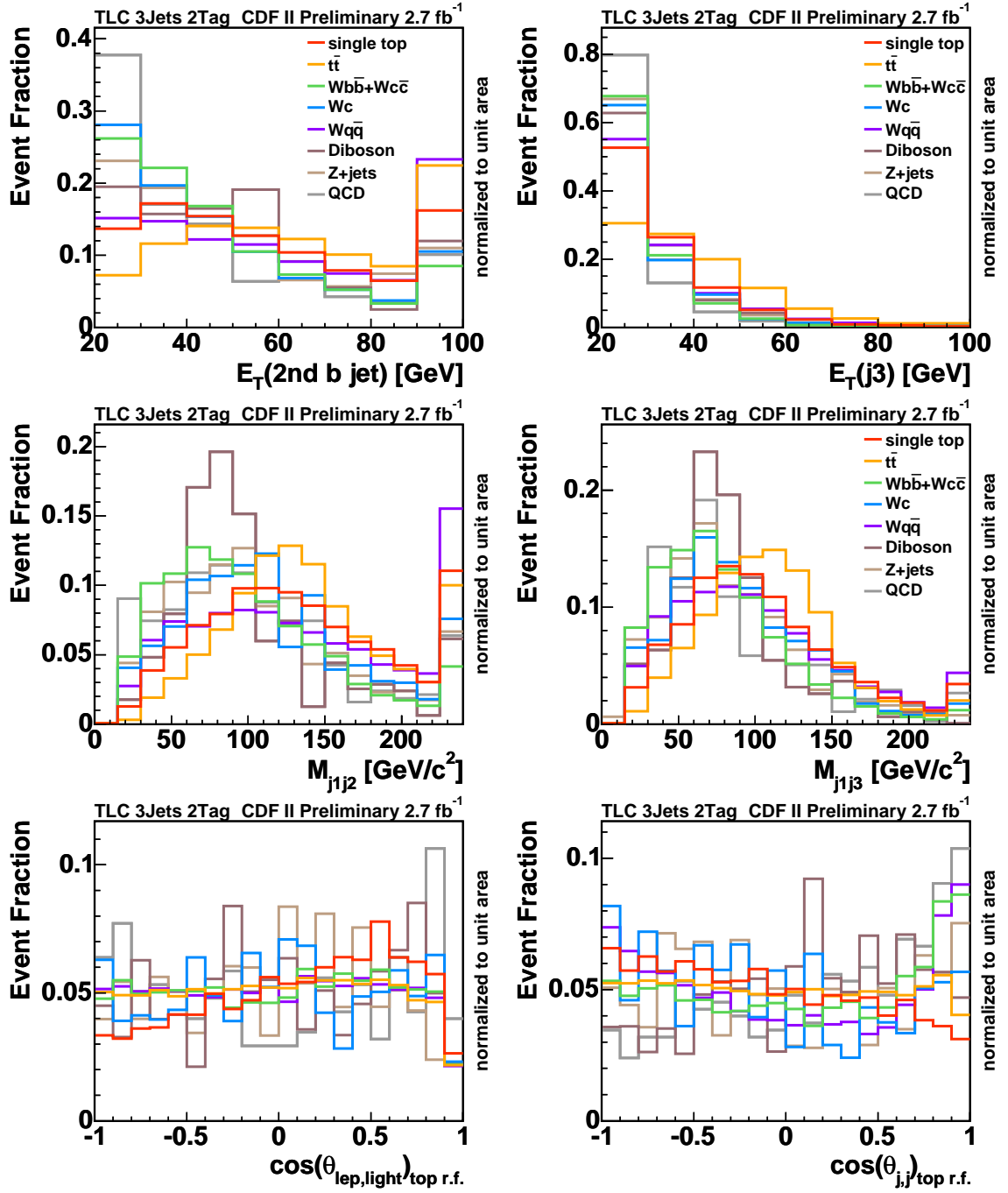


Figure 67: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for TLC leptons.

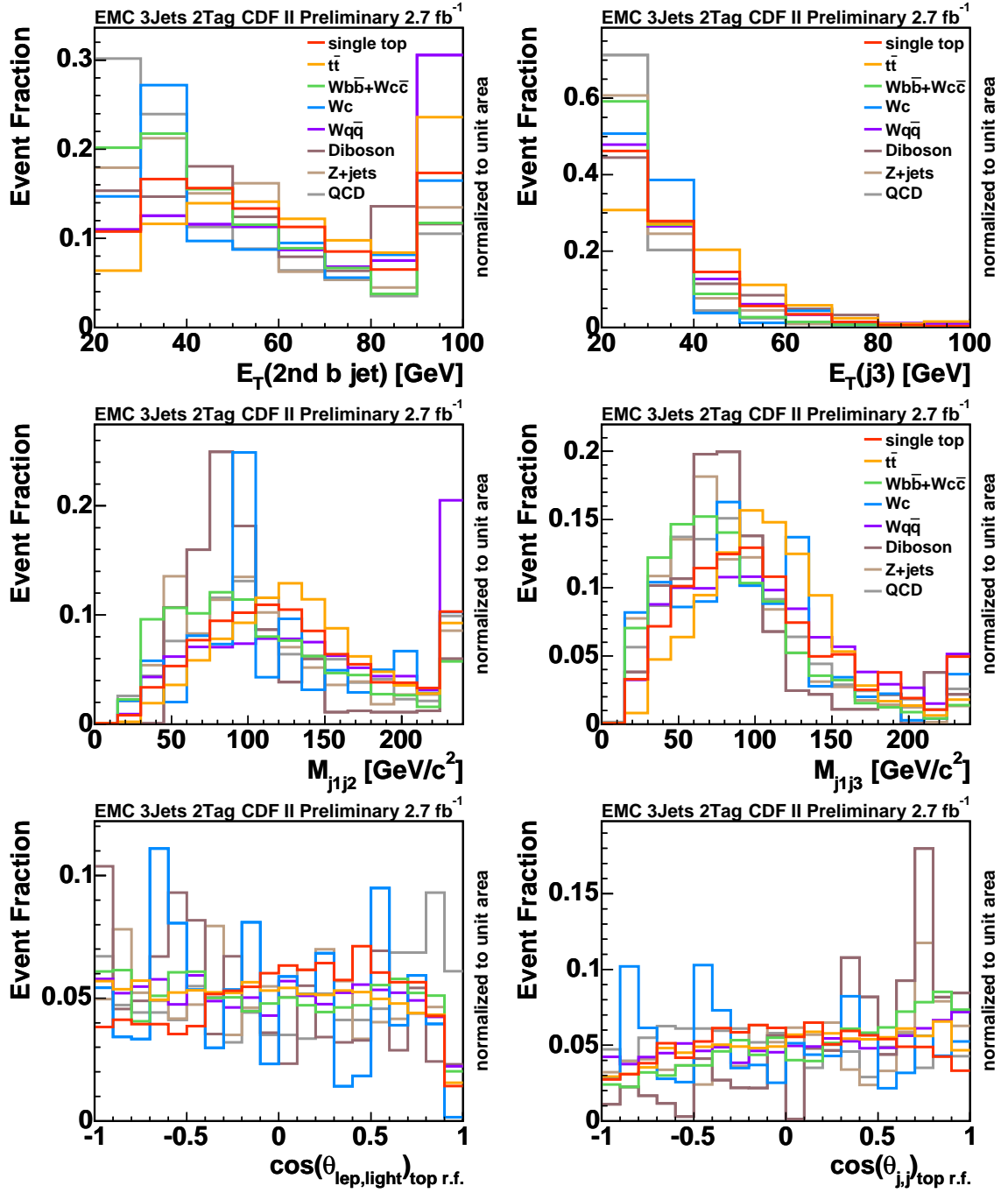


Figure 68: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for EMC leptons.

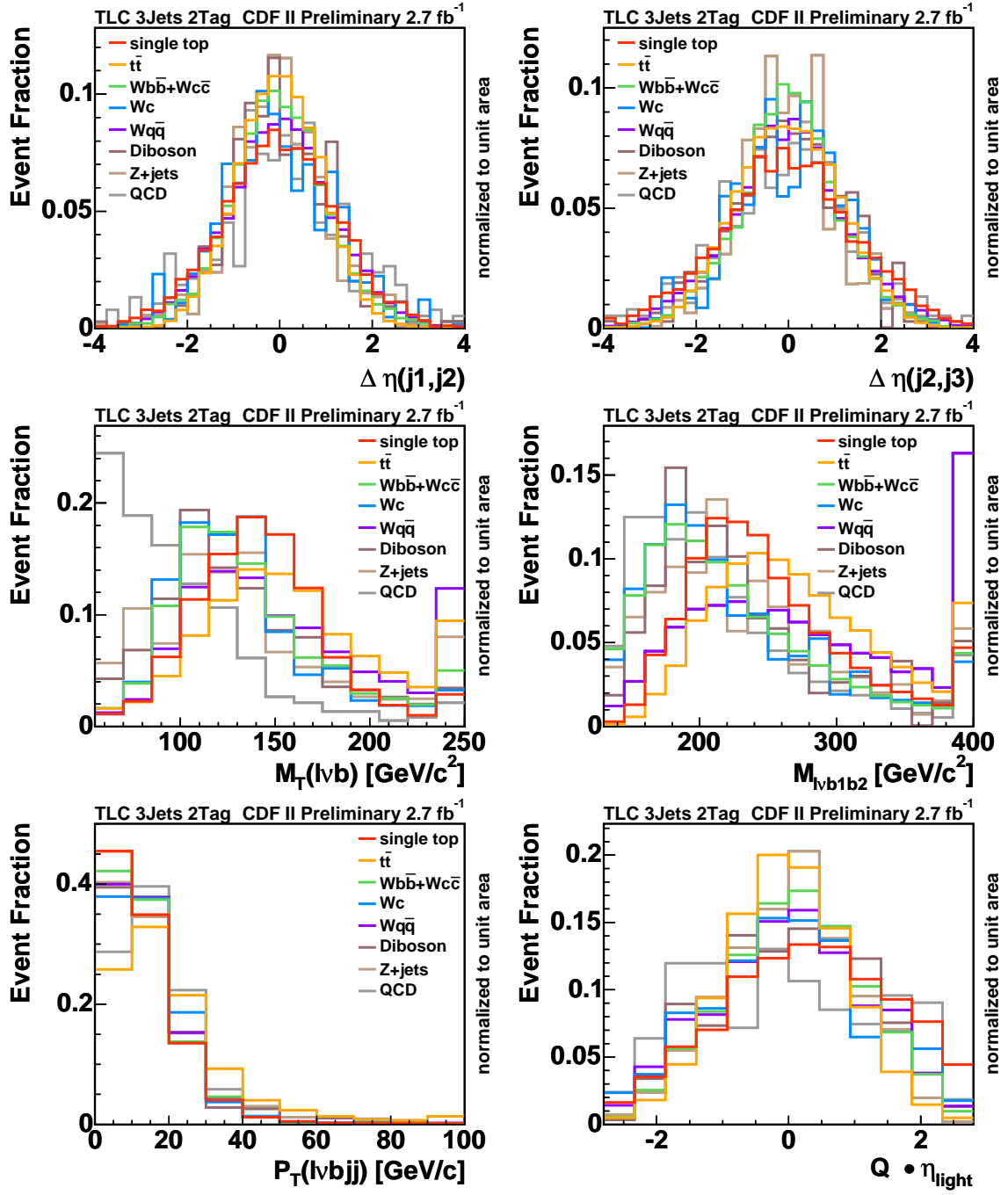


Figure 69: Shape comparison for input variables used in the t-channel neural network requiering 3 jets and 2 tags for TLC leptons.

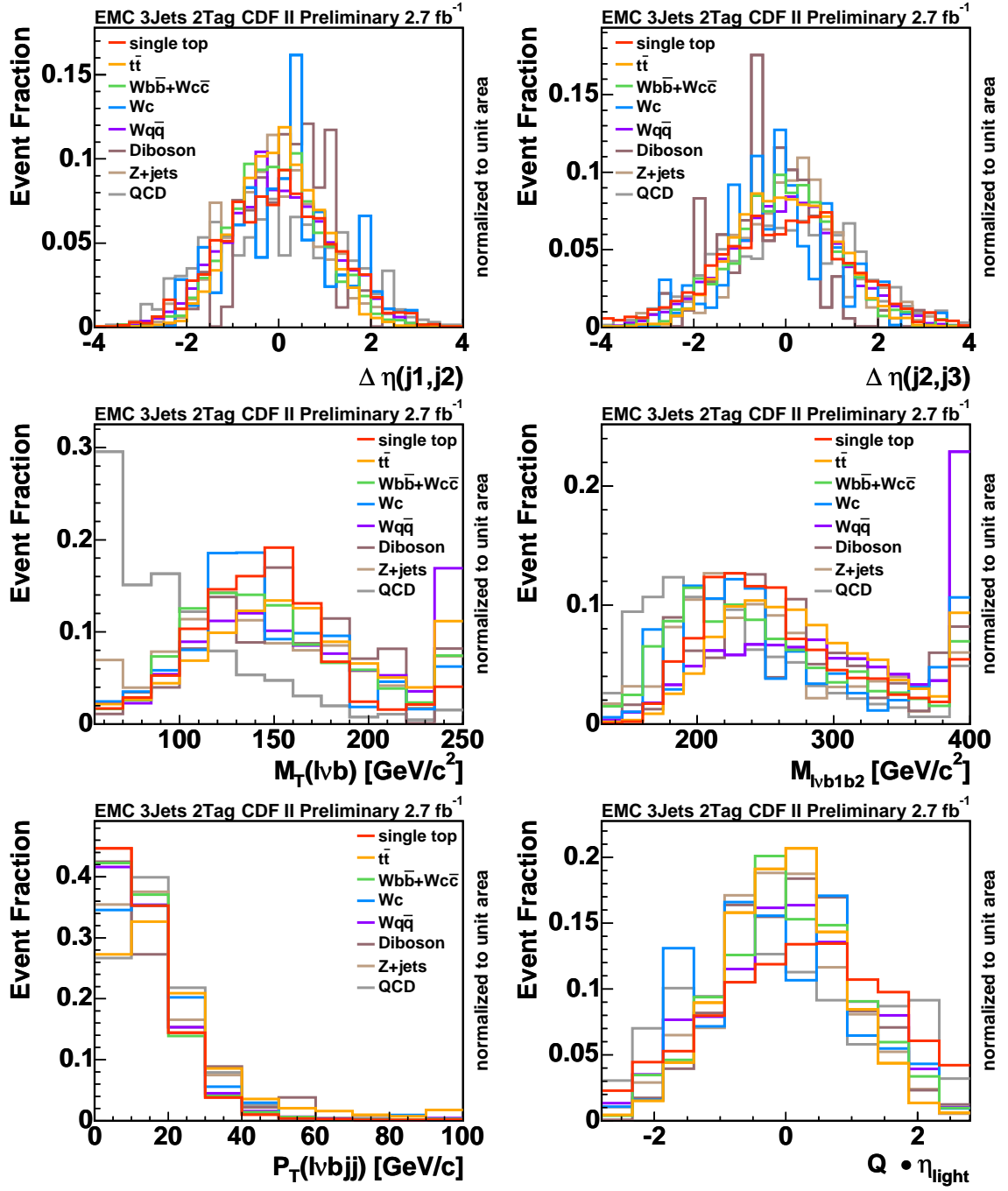


Figure 70: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for EMC leptons.

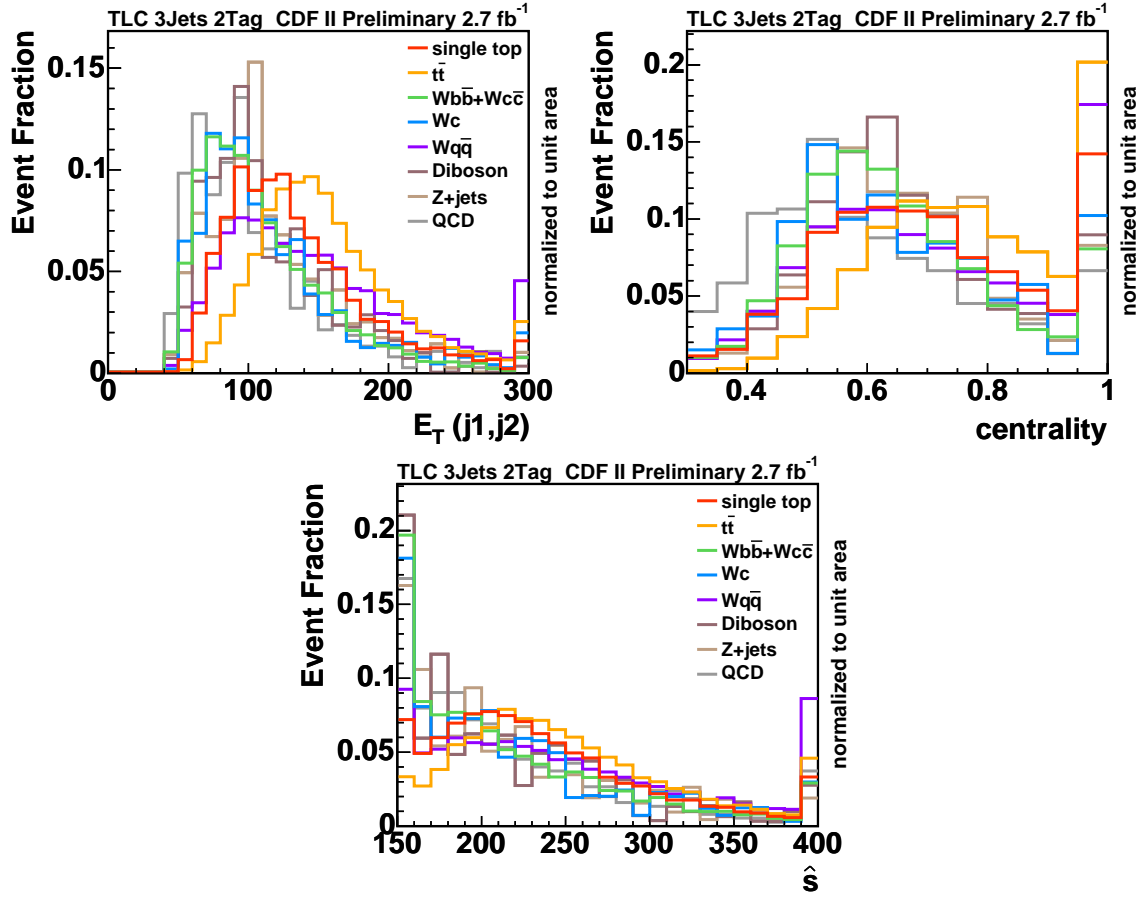


Figure 71: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for TLC leptons.

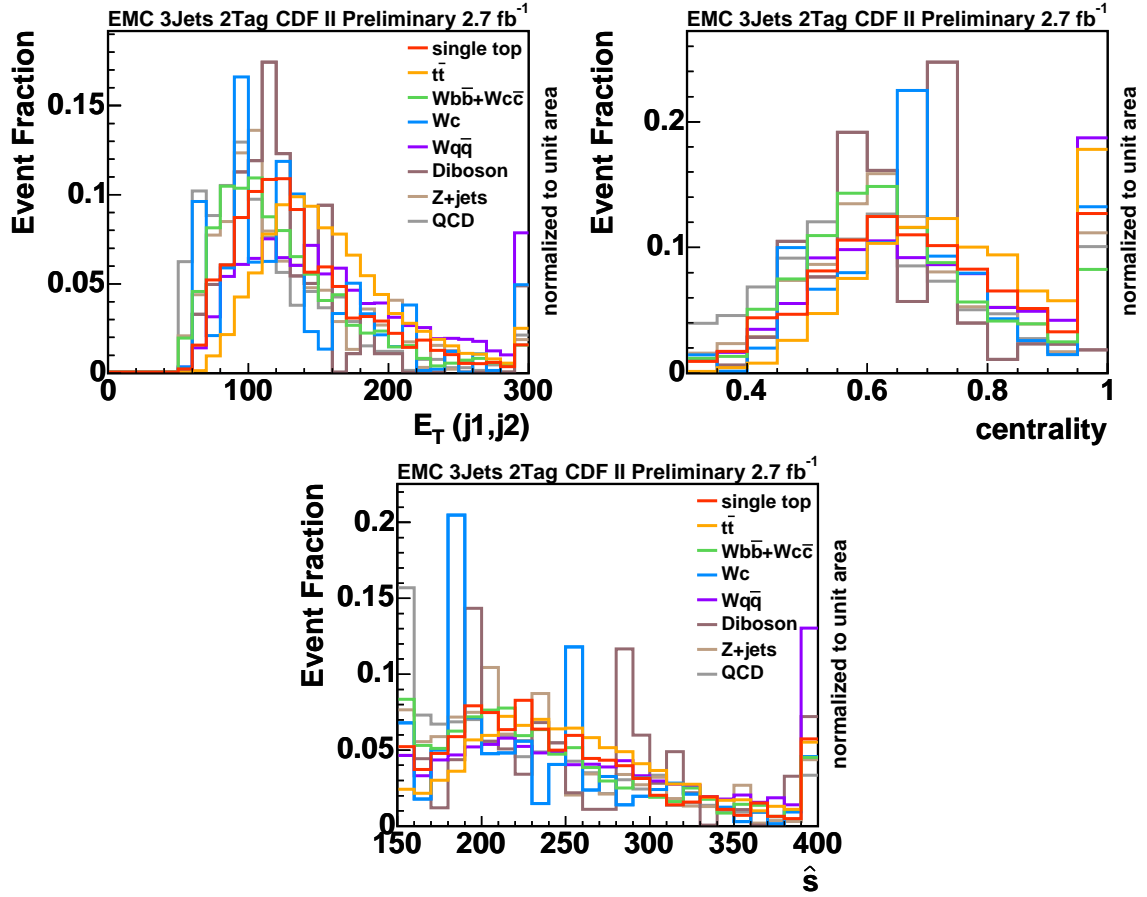


Figure 72: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for EMC leptons.

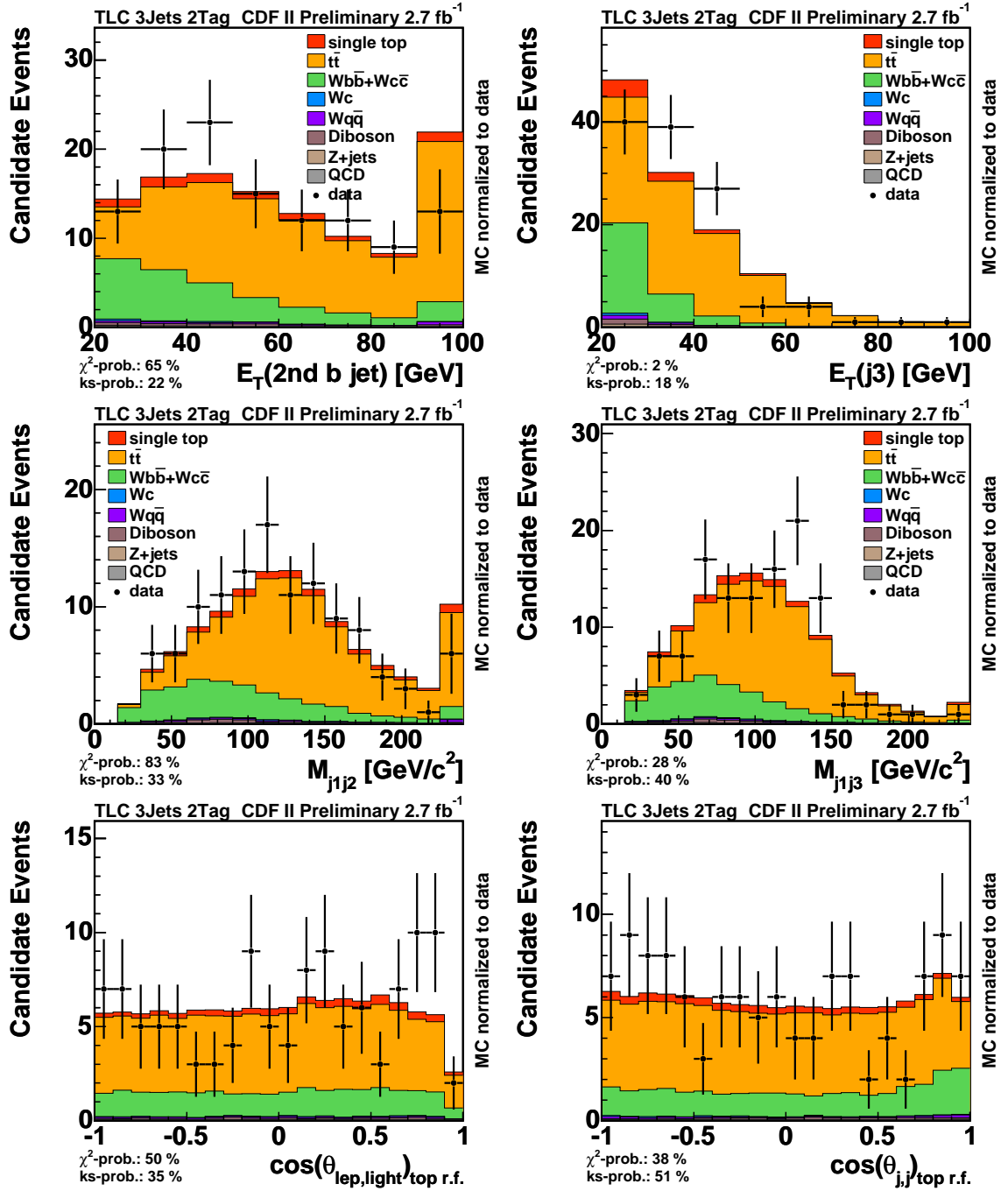


Figure 73: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for TLC leptons.

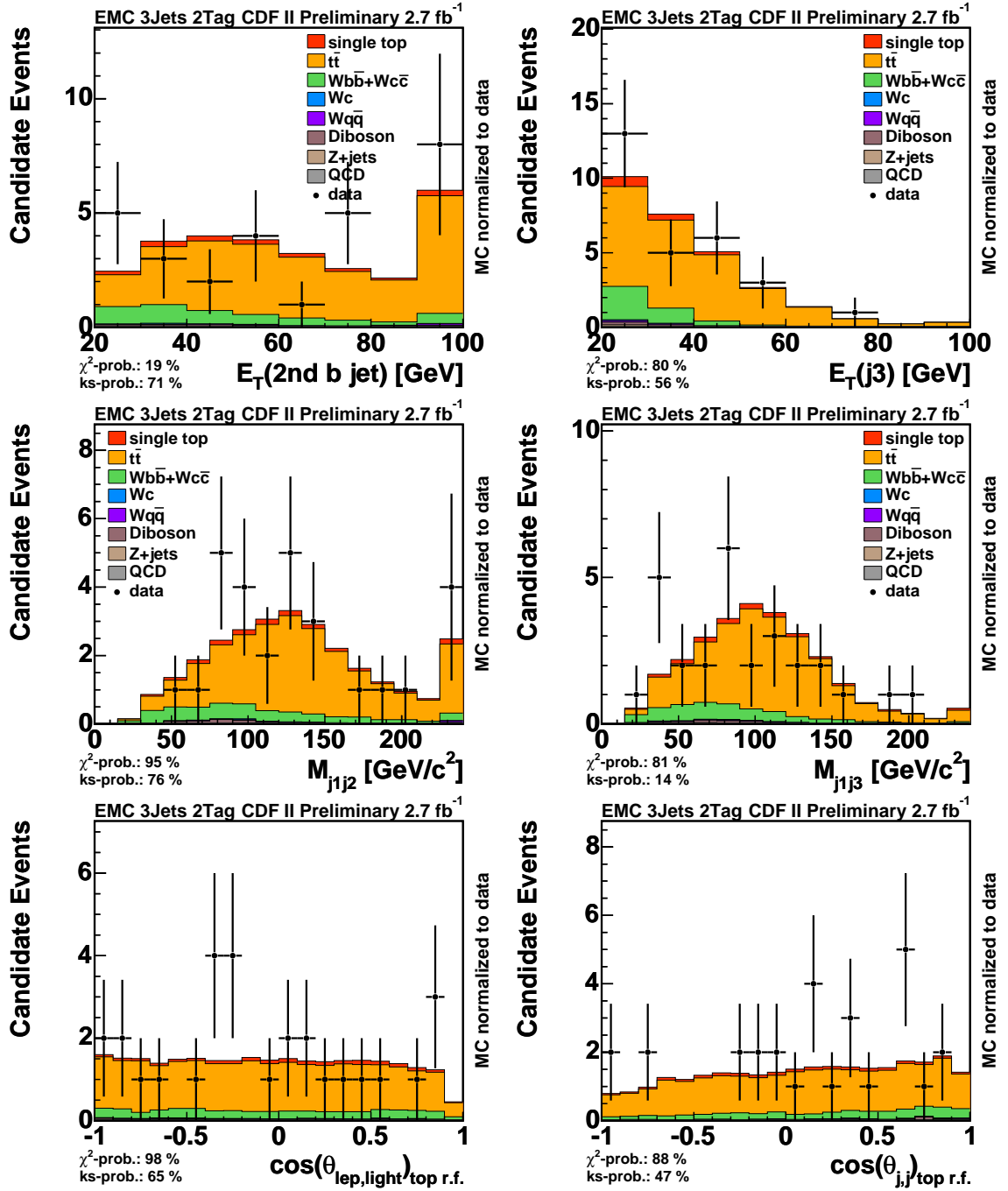


Figure 74: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for EMC leptons.

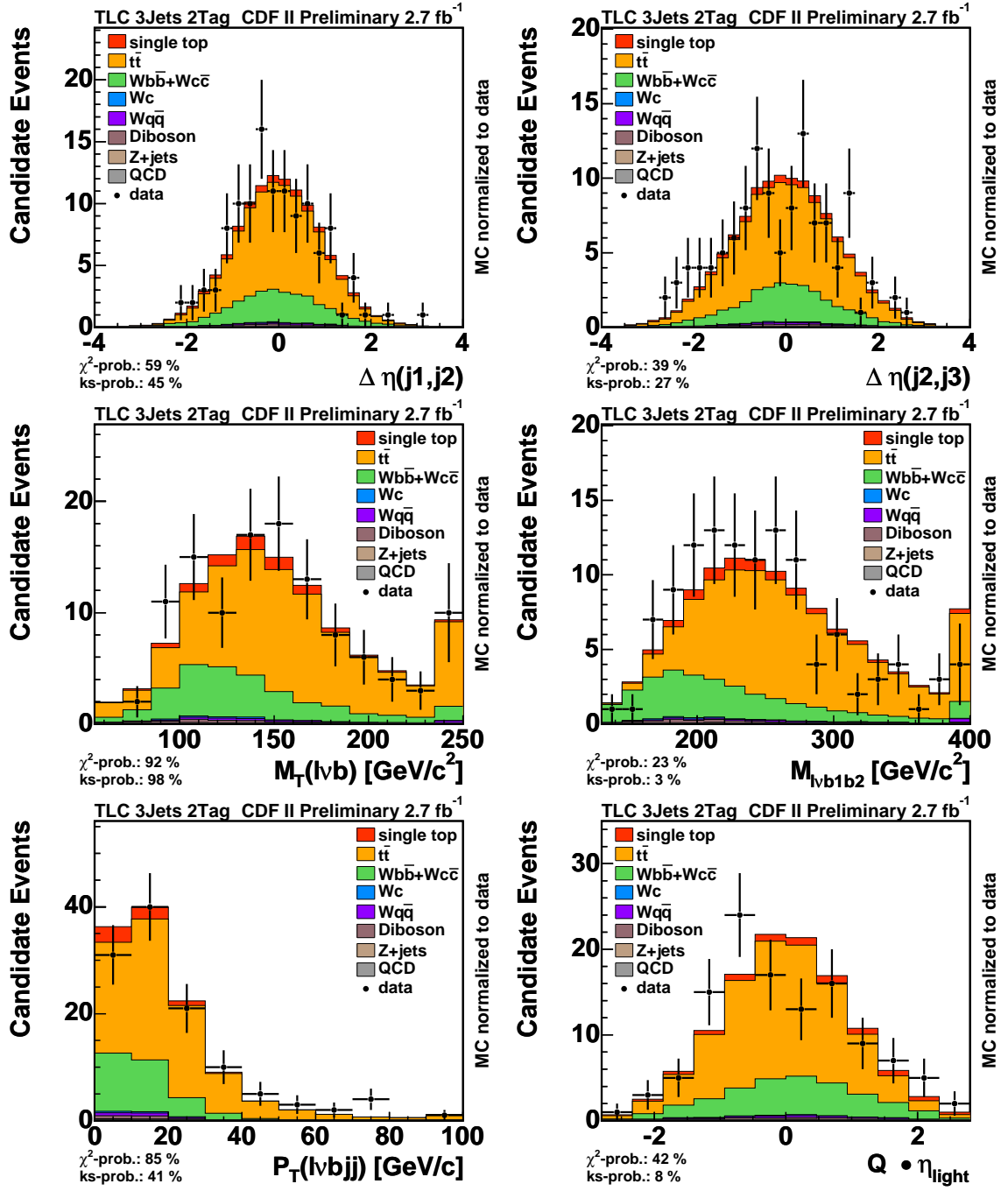


Figure 75: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for TLC leptons.

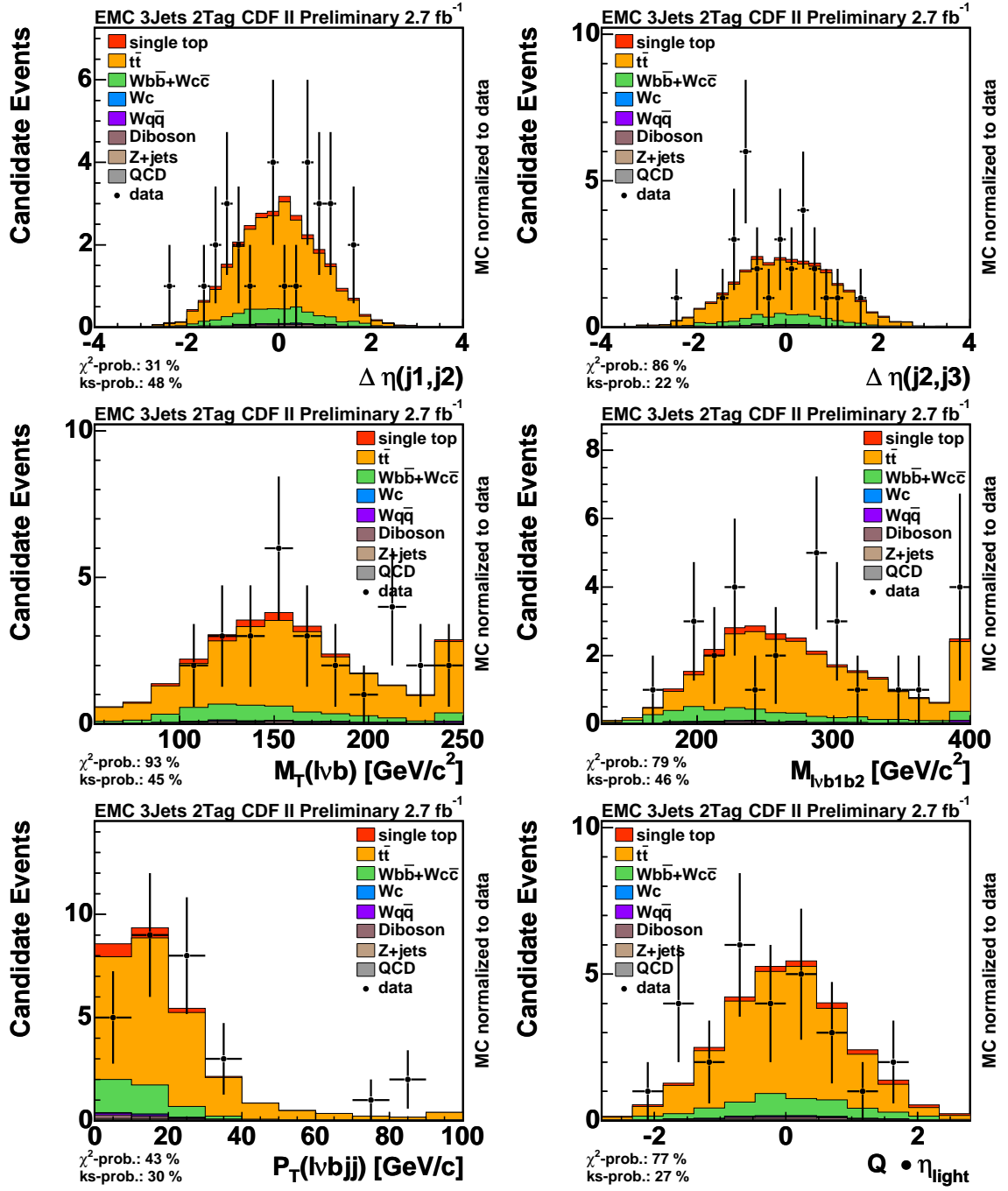


Figure 76: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for EMC leptons.

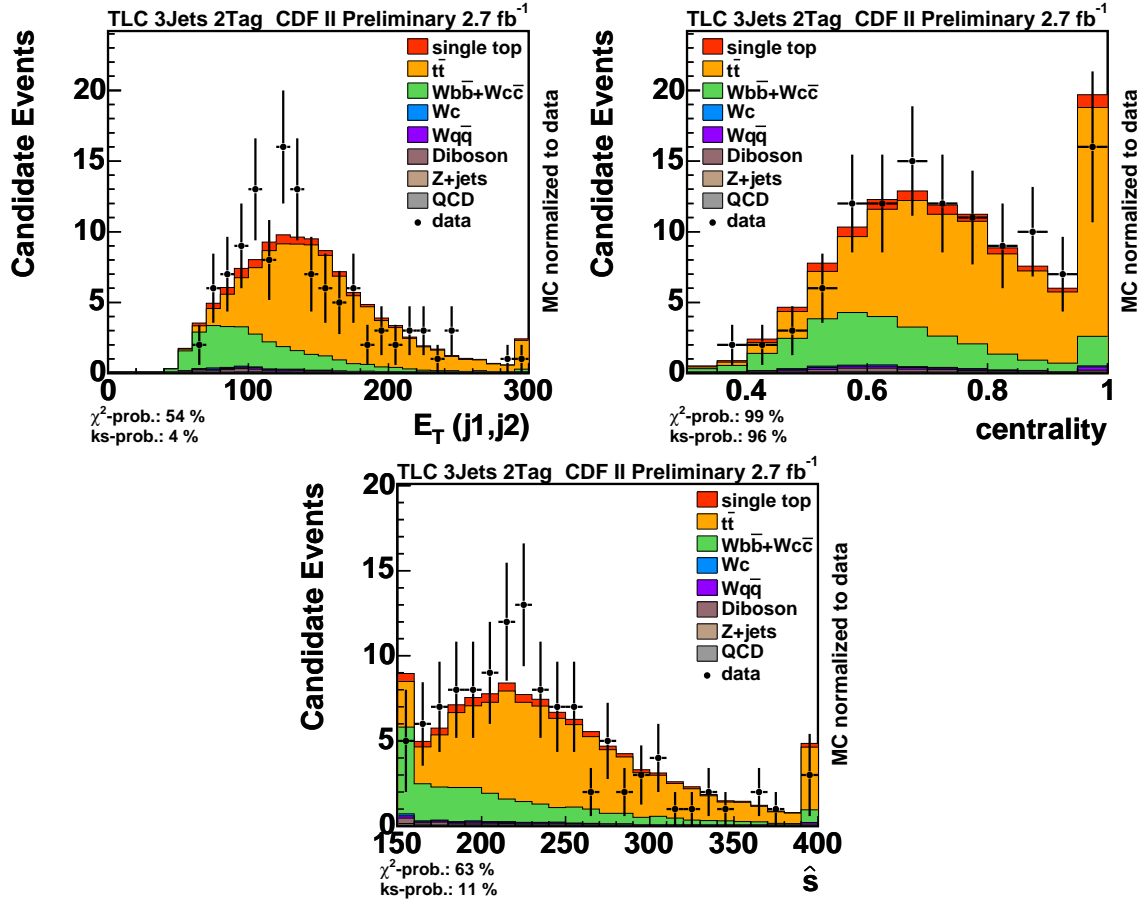


Figure 77: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for TLC leptons.

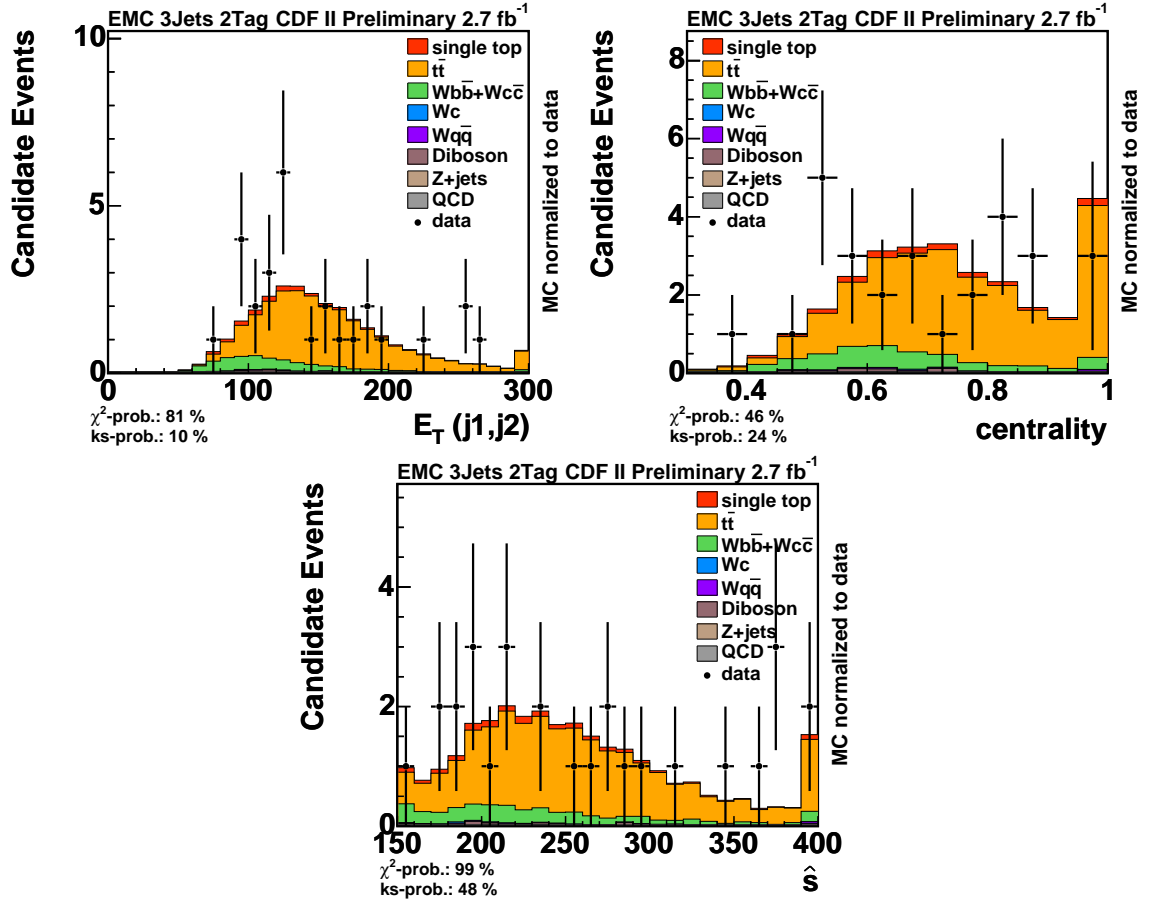


Figure 78: Shape comparison for input variables used in the t-channel neural network requiring 3 jets and 2 tags for EMC leptons.